

Is the world running out of wealth?

This is a bad time for Mr Mikhail Gorbachev to be looking for assistance from the industrialized West, but he should not be sent away empty-handed, while the West should plan now to avoid the next recession.

SOME parts of the industrialized world seem to be heading for needless trouble. In the United States, Wall Street and the administration are repeatedly seeing signs that the recession is nearing its end — and, so far, are repeatedly disappointed. Elsewhere, in Britain for example, what the government used to hope (just two years ago) would be a “soft landing” (from the heights of the 1980s boom) has become a bumpy one. The government no longer says that the bottom of the recession is round the corner, but that it will be “in the second half of the year”; by then, more businesses will have collapsed, and unemployed people will number close on three million. And now the trouble seems to be spreading to France. What has gone wrong?

On the face of things, there is no rhyme or reason for these happenings. If the adult members of a population of 5,000 million were engaged on making motor-cars in modern plants, each of them would be able to change his or her motor car every week or so. If the same adults were employed in modern agriculture, the European Communities’ surpluses of food would be multiplied gigantically. If the same people were engaged on the mix of activities that occupy Japan’s population of 125 million, the world collectively would be forty times richer than Japan. So why not simply arrange for that or, alternatively, get rid of the present governments of industrialized countries and replace them with another crew?

The trouble is that matters are not nearly as simply arranged. Making motor-cars as efficiently as they are made in Japan, or growing cereals as efficiently as in the Middle West of the United States (this year’s droughts, floods and tornadoes notwithstanding) requires the investment of large amounts of capital. And capital investment, as Marx knew, cannot be secured by printing banknotes, but requires that people should forego consumption in the present so that others can be paid to win the natural resources and make the machines that will eventually allow production to be as efficient as it can be — by demonstration — and more (which is the promise of research and development).

The present recession in a handful of industrialized countries (so far) has arisen because the consumption and investment were not well synchronized as the 1980s drew to a close. The United States was outrageously the worst offender, indirectly financing a consistently high budget deficit by the foreign purchase of domestic assets and failing to restrain consumption to the extent that savings would pay for the creation of future assets. Much the same happened in Britain except that the government ran not a deficit but a surplus —

at the expense of investment in the infrastructure of an industrial country.

Quite what mixture of factors, last year, precipitated the recession remains to be determined, but the loss of wealth (or the sense thereof) in the stock-market crash four years ago and the commercial banks’ succession of disasters over loans to developing countries will be found to have played a part, as will be the general overexpansion of credit that followed the crash. What has been happening since is that the economies chiefly affected have been compelled to seek competitiveness by shedding marginally uncompetitive enterprises, in the process making worthless much recent real investment, many claims on future consumption (such as the value of pension funds) and putting people out of work.

These are hardly propitious circumstances in which Mr Mikhail Gorbachev, on behalf of the Soviet Union, should be appealing to the West for help. But in the same language, his problem is more serious than anybody’s — except for the heads of the governments of the developing countries for whom the hope of relief from poverty seems beyond grasp. The programme of productive investment on which Lenin embarked and Stalin continued was based on the calculation that enforced savings in a managed economy would yield prosperity. But the investments were badly managed and proved unproductive. Now the Soviet people have had enough, and even Mr Gorbachev could not persuade them to embark on another seventy years of enforced saving on behalf of their grandchildren’s comfort.

What that means is that the Soviet Union must look to others to make the machines and other capital goods that might eventually make the Soviet Union productive. That, in essence, is what he said in his Nobel Peace Prize address last week in Oslo, and what he is likely to repeat at the economic summit in London next month. But although the recession should — on past form — have induced people to save (against the prospect that there may be worse to come) even when the recession has ended, the temptation will be to invest at home, where unemployment has risen, rather than on some distant steppe. That, at least, is what the West seems tempted to say.

The more imaginative response is more risky, but not impossibly so. Mr Gorbachev may be asking for a lot of money, but he is not asking for the Moon. By some accounts, \$1,000,000 million (the Pentagon’s annual spending) over several years would suit him, by others, \$250,000 a year (a little larger than the annual US federal deficit) for several



years would do. The numbers may be large, but they are not unthinkable. But if funds are to flow on this scale to the Soviet Union, they can come on this scale only from private sources — pension funds, for example — and will have to be raised in competition with less obviously risky investments in the West.

Can Mr Gorbachev promise security of that kind, and then deliver? Or could he and his constituents settle for the more galling alternatives of investment exclusively in dissident republics (such as the Baltic states), which would seem like external interference but which would be none the less worthwhile, or for investment in one industry (such as petroleum) exclusively, which would be the easiest way of revitalizing the Soviet economy and its export business, but which would smack of neocolonialism? Those are the questions Mr Gorbachev will be faced with next month.

Meanwhile, the industrialized countries of the West (perhaps not Japan) will be faced with other conundrums, of which the chief must be to reach a general understanding of how future recessions should be avoided. The imbalances of the past few years have mostly been brought about not so much by deficiencies inherent in market economics, but by unreasonable expectations of what markets can accomplish. Since the aggregate value of a community's consumption must in the end be related to the aggregate of the physical resources (including capital goods) that it uses up, people can become richer by buying and selling pieces of paper such as stock certificates or deeds of title to land or houses only at the expense of others. Governments, this bad patch may suggest, have a duty to exert an influence in that direction, perhaps by increasing taxes in good times and decreasing them in bad — the Keynesian recipe. Sadly, in the United States and Britain, the opposite recipe has been followed for ten years. Is it any wonder that they are in the most trouble now? □

Electoral science

British Labour's plans for science are none the worse for their moderation.

It is an open secret that there must be a general election in Britain in the next year, although government spokesmen will only barely acknowledge this constitutional necessity. Meanwhile, both the government party and the opposition parties are in full cry, raising doubts about their stamina to keep things going at this pitch for very long. This week, the opposition Labour Party did much to justify the opinion that it has made itself "electable" (after three defeats in a row) with its policy for research and development (see page 512).

Despite the government party's earlier theft of some of Labour's clothes, as in the announcement two weeks ago that the government would also abolish the distinction between universities and polytechnics, the Labour document on science has put together a sensible policy. Perhaps more important, its plans are free from the excesses of enthusiasm for what Mr Harold Wilson (as he was in 1964) called "white-hot technology". The new recipe is none the worse

for having only dull red-heat.

The proposals for the administration of civil science are radical, but inoffensive. Most things follow from the decision to bite an old bullet and appoint a minister for science. The past few years have shown that there is a need (as the House of Lords has been urging for the whole decade), most obviously in the external representation of the British position but more seriously in the coordination of the government's own policy on science and technology.

Labour would also plan to offer tax credits for industrial spending on research and development, which is an improvement on its proposal at the last election in 1987 that a substantial increase of spending should be brought about by a levy on industrial earnings. Even foreign-owned companies would qualify for tax rebates. The objective is to increase civil research and development to 2.5 per cent of Gross Domestic Product which, the policy document says, is "still below ... Japan's level of spending". There is no talk of requiring civil servants to be good at "picking winners". It will be interesting to see how many of these proposals will be picked up by the Cabinet committee on science and technology of which the Prime Minister is the chairman. □

FDA and immortality

The new US Food and Drug Administration is fighting a good fight against advertising excesses.

THE educated public in the United States seems to be consumed by a desire not to consume fat. The expectation is that one will thereby live forever. Duly alert to the health craze sweeping the country, food manufacturers have been quick to label their products with the magic words "No cholesterol", often accompanied by a drawing of a Valentine's heart. "Fresh" has also been turning up on a lot of products.

But last November, David A. Kessler, with a Harvard MD and a law degree from the University of Chicago, became commissioner of the US Food and Drug Administration. Since then, lying labels have been disappearing from supermarket shelves. First, Kessler took on Procter & Gamble's Citrus Hill Fresh Choice orange juice, which is made from concentrate, not squeezed fresh from oranges at all. Kessler said the "fresh" had to go. It went.

Kessler also went after the "no cholesterol" labels on vegetable oils, arguing that even though they do not contain cholesterol, they are nonetheless 100 per cent fat. People buying them to avoid fat were being misled, the new commissioner observed. Once again, the company backed off.

Now, Kessler is attacking misleading fat labels across-the-board. Ice cream and frozen yogurt are on his list, for instance. When the commissioner has finished, the public will no longer be deceived. It is arguable whether this will make those people looking for an easy yet food-rich route to immortality happy; maybe some would secretly just as soon not know. Nevertheless, what Kessler is doing is good for the image of the FDA — an often beleaguered agency — and for the public, which cannot live forever on wishful thinking. □

Who needs a genome ethics treaty?

- Little consensus at international conference
- Could Germany's strict laws become model?

Bethesda, Maryland

Is the world ready for an international treaty to regulate the use of genome research? Not just yet, judging by a conference on the ethical and social issues raised by the human genome project, held at the National Institutes of Health (NIH) last week.

The conference, sponsored by the NIH National Center of Human Genome Research, brought bioethicists and policy-makers from around the world together with leaders from the US, Japanese and Soviet genome programmes. Unlike many previous NIH-funded genome ethics meetings, last week's gathering looked specifically at the need for an international genome ethics treaty. But with opinion sharply divided — largely along national lines — the group could resolve only to set up a series of international task groups to debate the issues further.

Peter Singer from Monash University in Australia, head of the newly formed International Association of Bioethics, questioned the goal of attaining a genome ethics convention signed by as many nations as possible. As long as there are large differences in ethical opinion from country to country, he said, an international treaty will achieve little, with vague wording reflecting only "a superficial level of agreement".

These differences in opinion are huge. For instance, the German Bundestag, sensitive to the suggestion of a revival of Nazi eugenics policies, voted last year to allow screening of embryos only for couples at risk of passing on severe sex-linked disorders, such as Duchenne muscular dystrophy (see *Nature* 348, 8; 1990).

Germany has also banned any attempts at human germ-line manipulation, by a unanimous vote of the Bundestag, says Michael Catenhusen, an opposition Social Democrat and chairman of the Bundestag's Science and Technology Committee. Feelings run so high that conferences in Germany to discuss bioethics have been called off, for fear of disruption by demonstrators from environmentalist, far-left and anarchist groups.

Asked about the prevailing attitude to genetic screening in Germany, James Watson, head of the NIH genome programme, said he could not see how a developed nation can continue being "backward", denying pre-implantation genetic screening for couples at risk of having children with genetic diseases, simply "by invoking the name of Hitler".

Catenhusen maintains that Germany's laws will become the model for the rest of the

world. But that seems unlikely — even within Europe, other countries have already passed far less restrictive laws governing genetic screening. And many of the geneticists attending last week's meeting were wary of strict legislative controls, at least so far as research is concerned. Lev Kisselev, from the Soviet Academy of Sciences in Moscow, hardened by the manipulation of Soviet biology under Lysenko said he was opposed to any interference with research by "non-competent" people.

So far, the 25-nation Council of Europe, a body strongly concerned with human rights issues, is the only international organization planning to negotiate a bioethics treaty. The Council of Europe convention will contain a number of protocols, each dealing with a specific area — such as medical research and the trade in body organs for transplantation. Countries will be able to ratify the convention à la carte, choosing to obey whichever protocols they choose.

But there are no plans yet for a protocol on ethical aspects of the genome project, and Frits Hondius, deputy director for legal affairs at the Council of Europe, is not convinced that there is a strong need for one. In any case, the Council of Europe may not be the best forum for an international genome ethics treaty when the two major players in genome research are the United States and Japan.

The task groups set up by last week's meeting will be looking at a number of key issues, including the use of genetic information to discriminate against individuals in employment and when arranging insurance, and the protection of genetic data stored in forensic databases. The task groups will eventually report to a steering committee made up of experts from each of the major committees around the world now struggling with genome ethics — including those from the European Communities, the Human Genome Organization, the Council of Europe, and the US, Soviet and Japanese genome programmes.

This new steering committee, which will itself tackle the divisive issue of the clinical uses of genome research, is seen by many delegates to last week's NIH meeting as the most positive development. Even if no international treaties are drafted, more dialogue between the various committees will help to prevent them duplicating each other's work and "reinventing the wheel", says Nancy Wexler, who chairs the joint NIH/Department of Energy genome ethics working group.

Peter Aldhous

NASA BUDGET

Space station back on track

Washington

THE threat to cancel the US-led space station Freedom (see *Nature* 351, 428; 6 June 1991) has subsided, at least for the time being, after the full House of Representatives voted last Thursday to provide \$1,900 million for the station in the 1992 fiscal year. The decision reversed that of the House Appropriations Committee, which had included only \$100 million for Freedom in 1992, barely enough to close the project down.

The House decision, passed by 240 votes to 173, would freeze the overall 1992 budget of the National Aeronautics and Space Administration (NASA) at its 1991 level. This should keep the space station project on course, but could delay science missions such as the Cassini Saturn probe and the CRAF cometary probe (which for budgetary purposes are packaged as a single mission), due for launch in 1995 and 1996 respectively, and the \$40,000 million Earth Observing System remote sensing programme. Spending on these two projects was scheduled to accelerate during the coming year.

In the case of CRAF/Cassini, where the 1992 budget request of \$328 million was more than twice that of 1991, the proposed budget freeze could even force cancellation, says Dale Cruikshank, chairman of the American Astronomical Society's Division of Planetary Sciences. There are only a limited number of dates on which the two probes can be launched to achieve their missions, and long delays will increase costs drastically.

The battle over NASA's budget now moves to the Senate. Last year, the Senate restored some threatened cuts to NASA's budget, in a compromise worked out between the House and Senate, which saved CRAF/Cassini. The Senate is expected to be tougher this year, however, so it is unlikely that all of NASA's proposed space science projects will be funded in full.

Last week's endorsement of the space station in the House followed intensive lobbying to save Freedom by President George Bush's administration. Bush himself is said to have telephoned key members of Congress to canvass their support, and Richard Darman, director of the White House Office of Management and Budget, told the pro-station House Committee on Science, Space and Technology last week that killing Freedom would not increase spending on other science projects. On previous occasions, when Congress cut money from the President's budget request for the Superconducting Super Collider, he said, the funds were reallocated to water projects, not science. Darman also said that Bush's senior aides would advise the President to veto any congressional attempt to cancel the space station. **Peter Aldhous**

"Too little, too late" for AZT

Washington

A FLURRY of recent legal challenges to Burroughs Wellcome's exclusive patent for the AIDS drug AZT may eventually wrest control of the patent from the pharmaceutical company, but the efforts have come too late significantly to lower the price of drug in the next year or two, when it is needed most, AIDS activists say.

Three challenges in the past several months have fuelled speculation that Wellcome will soon have to concede that US government and university scientists had a pivotal role in discovering and developing AZT and thus deserve part-ownership of the patent rights. Lawsuits by the generic drug company Barr Laboratories and the consumer group Public Citizen, on behalf of the People With AIDS Health Group and several AIDS patients, are seeking to overturn the company's patent. The National Institutes of Health (NIH), whose scientists first discovered that AZT is effective against AIDS in 1985, last month released a statement criticizing Wellcome for taking undue credit for the drug.

Although NIH have not yet decided to sue on their own behalf, agency lawyers are considering joining one of the other suits as a supporting player.

Although some of the challenges may eventually meet with success, activists say that none appear likely to change the pricing picture for AZT in the near term. The Public Citizen suit is likely to fail on "standing" grounds — consumers are usually not considered eligible to challenge a patent. Barr's suit has no standing problem, but, like any patent challenge, it is likely to take years to decide. "We're looking at a legal process so long," warns Derek Hodel of the PWA Health Group, "that the issue will be moot by the time it's settled." The NIH statement, meanwhile, coming nearly two years after government researchers first publicly criticized Wellcome for taking credit for AZT, is probably "too little and too late" to influence matters much now, says Jeff Levi, director of government affairs for the AIDS Action Council. He notes that although NIH lawyers have been considering legal action since last year, they have so far made no effort to follow through with an actual lawsuit. Last month's statement, by NIH director Bernadine Healy, is little more than an official endorsement of a dissenting position that NIH scientists have taken since the government granted Wellcome a patent for the drug in 1988.

Rather than holding out hope for success in the courts, AIDS activists believe that industry competition — in the form of new drugs — will be the factor that finally cuts the costs of AZT.

This fall, the Food and Drug Agency is expected to approve DDI and DDC, two new drugs for the treatment of AIDS.

Although initial clinical trial results suggest that neither drug is significantly better than AZT, the simple presence of three drugs from three different pharmaceutical companies will probably trigger a drop in prices. AZT currently costs between \$2,200 and \$2,800 for a yearly supply.

Last year, two Canadian drug companies decided to challenge the Wellcome monopoly by manufacturing AZT and selling it through outlets in the Bahamas, where Well-

HIGHER EDUCATION

British move sows confusion

London

THE ceremonial opening last Friday by the Princess Royal of a new building at Bristol to house the funding councils for both British universities and polytechnics may have turned out to be symbolic of the government's plan to merge them, but most other questions provoked by last month's announcement (see *Nature* 351, 257; 1991) have not yet been decided.

The overriding problem is that most of the planned changes will require legislation, and that the government must call a general election in not less than 12 months from now. Even if legislation were introduced in the parliamentary session due to begin next November, there might not be time for its enactment.

The government's intention seems to be to introduce in November legislation to allow for the transfer of further education colleges from local authority control to that of funding councils yet to be formed, as well as for its system of proposed National Vocational Qualifications. The proposals affecting universities and polytechnics would be introduced only after the next election (when the present government may no longer be in office).

Even the planned merger of the two funding councils, themselves established by the Education Reform Act of 1988, is in limbo, the new building notwithstanding. A spokesman for the Department of Education and Science said this week that the transitional arrangements had not yet been worked out. The government's intention is that the new funding council should be operating from 1 April 1994.

The appointment of Lord Chilver as chairman of the Universities Funding Council (UFC) terminates in November, but no decision has yet been made about his continuation even as a caretaker.

The notion that the merged funding council should then be divided into national entities is a further source of confusion. The planned separate funding council for Scotland is regarded even by academics at English universities as inevitable.

But there is anxiety about the intended

come does not have a patent (see *Nature* 349, 93; 10 January 1991). They are charging about two-thirds of the Wellcome price. But the move has had little effect on US users because so far no one has been willing to bring the drug into the United States for resale. "We have no plans to import AZT at the present," says Hodel. "Part of our purpose in filing our suit is that the Burroughs patent infringement suit [against the two Canadian companies] prevents it." But without the legal clout of NIH, "there's not much light at the end of the tunnel," he says.

Christopher Anderson & Diane Gershon

separate funding council for Wales, which includes one federated university and one polytechnic. There are also a further six higher education institutions now controlled by local education authorities which are due to be taken over by the Welsh Office from 1 April next year, together with responsibility for the Welsh polytechnic.

Both in Scotland and Wales, there are fears that continuing support for academic institutions is unlikely to match the support they now receive for research from the research councils. In Wales, the small size of the proposed system is a further cause for concern.

But the most common anxiety among British academics centres on research funds. The government is at present working on a scheme by which some funds would be transferred from the UFC to the research councils, which would then (from September 1992) pay directly attributable overhead costs (but not researchers' salaries) to grant recipients.

The Department of Education and Science will announce in the autumn its response to a proposal by the Advisory Board for the Research Councils that the sum transferred should be more like £150 million than the £100 million already announced.

The most serious fears among university researchers stem from their knowledge that the enhanced status of the polytechnics, will lead them to expect a greater proportion of the available research funds, so that the funds available will be more thinly spread.

Meanwhile, while awaiting the promised legislation, polytechnics are having fun choosing names for the universities they will be allowed to be under the new regime. The Polytechnic of Central London seems determined to be the University of Westminster, many polytechnics in cities with a university already (such as Leeds, Leicester and Nottingham) may become the "city university" thereof, while the Oxford Polytechnic may become the "new university" of the same name.

John Maddox

■ Labour Party pre-empted on polytechnics: see page 512.

ELECTRICAL GENERATION

Storage opens up options

Washington

EARLY this month, a small electric company in southern Alabama took a big step towards the future of the electric power industry. The Alabama Electric Cooperative in Andalusia opened a 110-MW storage plant that uses compressed air to store power during times of low demand for use during peak hours.

The plant, which is the first of its kind in the United States and only the second such in the world, is remarkable in its own right but perhaps even more notable for what it represents: an emerging trend in the electric utility industry to use energy storage instead of added capacity to meet increased demand. A combination of environmental concerns and economic and regulatory pressures are pushing utilities to operate more and more efficiently, and storage is certain to become an increasingly important part of electric companies' supply strategy.

The Andalusia storage plant is built above a salt cavern. During off-peak hours, such as night-time, relatively inexpensive electricity is used to pump air into the cavern. Then in periods of high demand, the compressed air is let out of the cavern to drive a turbine and generate power. To increase the efficiency of generation, the escaping air is heated with gas or fuel oil before it passes through the turbine.

Alabama Electric decided upon compressed air storage after a detailed economic analysis, says John Howard, vice president of power production. "About 80 to 85 per cent of our generation was in base-load, coal-fired plants, so we really needed some peaking capacity," Howard says. Coal- and oil-burning plants as well as nuclear plants

are relatively expensive to build and relatively inexpensive to operate, so they are economical only if they are run nearly full-time. The compressed-air system, which is less costly to construct and more expensive to operate (per kilowatt), is a better option for providing electricity four to eight hours a day.

The compressed-air technology has been under development for more than 15 years,



Chino battery storage plant

and a plant in Germany has been using an early version since 1979, but the US electric power industry has been slow to warm to it. Much of the reason is simply that most US companies have more capacity than they need. "The whole industry is just beginning to come out of a glut of baseline capacity brought on in the 1970s," Howard notes. But rising electric use now means that utilities must find ways to handle ever-growing peak demands, and adding more baseline capacity is not the answer.

The Alabama cooperative settled on a compressed-air system because it had a

nearby salt dome that had been mined for many years. This not only offered it an easy-to-finish reservoir but also enabled the company to get regulatory approval for the system much faster, as there were few new environmental impacts to consider.

Utilities that have no nearby salt caverns have other options, says James Birk of the Electric Power Research Institute in Palo Alto, California. Many companies use pumped hydroelectric storage, for instance, but such plants can be located only in areas with the necessary bodies of water.

Battery storage is just beginning to be economically feasible. Southern California Edison has been operating a 10MW battery storage facility for more than three years. The plant near Chino uses lead-acid batteries like those in automobiles to smooth out load demands on its generating plants as well as to control the frequency and voltage of the electricity supplied to its customers.

The lead-acid batteries are expensive, however, and battery storage will not become widespread until more efficient, longer-lived, and less expensive batteries are available.

Researchers with the Electric Power Research Institute and other agencies are now testing such batteries as sodium sulphur and zinc-bromide to find an acceptable successor to lead-acid.

Another storage option is the use of superconducting coils, in which electric current can circulate without loss to resistivity. Superconductivity Inc. in Madison, Wisconsin, already offers small superconducting storage devices that can provide 1 MW of power for about one second. The devices are sold to industrial customers who use them to get rid of small voltage fluctuations that can disrupt sensitive equipment.

With the discovery of high-temperature superconductors, many people predicted that large superconducting magnetic energy storage systems would become practical. At the moment, however, low technology options such as compressed air and batteries seem likely to be more economical.

Whatever the technology, electric storage will be an increasingly important tool for electric companies, Birk says. Storage gives utilities the same advantages that the hub-and-spoke systems of flights give airlines, he explains. Instead of trying to fly each customer directly from his point of origin to his destination, airlines increase efficiency by routing all flights through a few hubs. Similarly, electric utilities can increase efficiency by not trying to move all electricity directly from a generating plant to a customer but instead by adding "hubs" — energy storage systems.

Once electricity providers learn to take advantage of the versatility and flexibility afforded by such storage, they will be able to provide better service with less generating capacity, Birk says.

Robert Pool

BEHAVIOURAL MEDICINE

Mental health research will return to NIH

Washington

IF everything goes according to plan, the National Institute of Mental Health will be moved out of the sprawling Alcohol, Drug Abuse, and Mental Health Administration and back to its previous home at the National Institutes of Health (NIH). The move, which is expected to be proposed formally by Bush Administration officials any day now, would symbolize a renewed emphasis on the basic and clinical research focus of mental health research.

According to a draft plan, the National Institute of Alcohol Abuse and the National Institute of Drug Addiction would also be moved into the NIH fold, marking a decision to put all basic research in behavioural medicine at the NIH. The current alcohol and drug abuse administration would deal exclusively with the service components of mental health care.

Discussions about whether to separate mental health research from health care

delivery have been going on for years — ever since the mental health institute was taken out of NIH more than 20 years ago, in fact — and the current proposal is likely to draw fire from mental health groups which will argue that their visibility will be diminished if they become part of NIH. But the proposal to consolidate all facets of mental health research appears now to have strong backing in the Administration and Congress.

Frederick Goodwin, the current administrator of the alcohol and drug abuse administration, has long favoured the separation of research from health care delivery and may even be a candidate for the vacant directorship of the mental health institute. The proposed merger is also favourably regarded by Bernadine Healy, the new director of the NIH, who has spoken out about the importance of research in neuroscience and its link to behavioural science. Barbara J. Culliton

Primate imports confusion

Washington

MORE than a year after a monkey virus related to the deadly Ebola strain infected four US animal handlers and shut down the primate import trade, state and federal health officials are still at a stalemate over regulations governing the importation of laboratory primates.

Despairing of a solution soon, one of the country's largest importers has abandoned New York state, whose stringent regulations are effectively prohibiting primate imports. Charles River Laboratories is moving its primate operations to an existing facility in Houston, Texas, which does not have the strict antibody test requirement demanded in New York. Other importers complain that federal regulators are dragging their heels on evaluating the problem, and that by allowing emergency measures to remain in effect for 16 months, the regulators have cost researchers millions in unnecessary primate expenses.

State and federal regulations have remained essentially unchanged since March last year, when the federal Centers for Disease Control (CDC) and several states imposed new "temporary" quarantine and testing procedures to ensure that imported monkeys are not infected with a filovirus related to the strain that killed hundreds of people in an Ebola outbreak in Zaire in 1976. Since then, researchers have found that the current virus is not Ebola, but another filovirus that is difficult to transmit and does not appear to threaten humans.

Nevertheless, the CDC as well as New York, California and other states have retained their emergency import restrictions. CDC require special airline handling procedures, and a 31-day quarantine coupled with blood titre tests at the beginning and end to ensure that the animals do not have an active infection. New York requires that the monkeys be serotested and quarantined in their country of origin and certified to be entirely free of any signs of filovirus infection.

Leo Grady, chief of the New York Department of Health virology laboratory, says health officials would be willing to lift the state restriction as soon as they see experimental evidence that latent filovirus infections cannot spread to other primates and humans. But such studies are not easy to do and only two US facilities — CDC laboratories in Atlanta and the Army's Medical Research Institute for Infectious Diseases — have the required resources.

CDC has done the experiment, says Kenneth Herrmann, deputy director of the agency's division of viral and rickettsial diseases. Two studies with 16 monkeys and two strains of virus resulted in "not one instance where the virus or viral antigen was found two weeks after infection," he says. "We've proved to our satisfaction that there really is no risk [of transmission] from a seropositive animal that is not ill." But the data will not be

published until fall, he says, and CDC has no plans to lift its own regulations until the data is out.

This goes down badly with researchers and primate importers, who were hit hardest by the emergency measures last year. After CDC temporarily shut down several of the largest primate importers and all US airlines decided to stop carrying the monkeys, the primate trade came to a virtual halt. Although CDC eventually recertified the importers, the continuing state and federal import restrictions and the forced dependence on non-US air carriers has more than doubled the cost of research primates, says Robert Doolin, director of Charles River Laboratories' national primate operations.

Cynomolgus monkeys that cost between \$300 and \$400 in 1989 now cost between \$1,500 and \$1,600, in part because transportation and handling costs have tripled.

In a letter of 9 May to CDC chief William Roper, Paul Schilling, director of Charles

FOOD AND DRUG ADMINISTRATION

Tempest over polio vaccine

Washington

DESPITE protests from the US pharmaceutical industry, government regulators are preparing to relax standards for the testing of polio vaccines to allow foreign competition in the \$100 million US market.

Last month, the Food and Drug Administration (FDA) published new regulations that would for the first time allow the US sale of polio vaccines that have passed the current World Health Organization (WHO) test instead of the usual US test. FDA hopes that some of the non-US companies that now sell WHO-tested vaccines elsewhere in the world will begin to market them in the United States, which would probably lower the price of the vaccine to government health agencies and to the public. On the basis of a five year review of worldwide data, the FDA concluded that both tests are equally effective in screening vaccine batches for safety.

But the one company that now sells the polio vaccine in the United States is fighting the decision. Lederle Laboratories, which has had the US market to itself for more than a decade, says that the FDA decision will force it from the polio vaccine market by putting it in an untenable product liability position. For the past five years, the company has argued that the epidemiological data that FDA uses may be flawed by under-reporting of polio cases and the low potency of some vaccines used outside of the United States.

Now that FDA has rejected those arguments, Lederle is worried that its strategy may have backfired. There are only about five cases of vaccine-induced polio in the United States each year, but, in almost every instance, the victim successfully sues

River's primate breeding operations and president of the Association of Primate Veterinarians, warned that the CDC regulations were costing the research community at least \$12 million a year.

Schilling estimates that because of the various restrictions, about 20 per cent of the nationwide primate demand last year went unfilled. Before the New York emergency regulations took effect last March, 80 per cent of the 20,000 research primates that were imported into the United States yearly came through John F. Kennedy Airport. Virtually no primates have come through JFK since then, says Grady.

Although researchers are hoping that publication of the CDC data this fall will put an end to the "temporary" nationwide restriction, New York may be an exception. No decision is expected there until a new state health commissioner is appointed to replace David Axelrod, who drove through the rigid regulations last March and has been in a coma for several months with a serious illness.

Christopher Anderson

Lederle. Million-dollar settlements are common. If FDA allows foreign companies to use the WHO test, Lederle would be put in an awkward legal position whenever someone comes down with the disease. If the company has not switched to the WHO test, a plaintiff could reasonably ask why it had not adopted the international standard that FDA had, after all, designated as "referred". If Lederle does use the WHO test, the victim could point out that the company itself had gone to great lengths over the past five years to show that the test may be unsafe.

So, to avoid an automatic court loss every time a vaccination goes awry, Lederle may have to use both tests, at more than twice the cost of the WHO test alone. That, the company argues, could put it out of the polio vaccine business because its competitors, using only one test, would have a cost advantage.

Lederle has at least one powerful friend in the matter. Late last month, Representative John Dingell (Democrat, Michigan), the chairman of the House Energy and Commerce Committee, warned FDA commissioner David Kessler that he would not stand for regulations that disadvantage US industry. "I am strongly opposed to any situation where regulation promulgated by a US regulatory agency put a US company which uses US standards at risk," Dingell wrote in a letter dated 21 May.

FDA officials have until the end of this month to decide how to respond. But unless the agency reverses its decision, WHO-tested polio vaccines will become legal in the United States on 21 June, and Lederle lawyers will have to begin preparing for the worst.

Christopher Anderson

JAPAN

Volcano prediction problems

Tokyo

LAST week's eruption of Mount Unzen, which killed 37 people, has highlighted the problems of volcano prediction and disaster prevention in Japan. In the days leading up to the Unzen tragedy, Japan's volcano experts and disaster prevention officials recognized the danger of the mountain erupting and even evacuated a nearby residential area, but they did not fully anticipate the particular threat that claimed the 37 lives. The episode is the latest in a string of recent volcanic eruptions in Japan that have defied the public predictions of experts.

Before the eruption, the warnings of various government disaster prevention organizations focused public attention on the possibility that mudflows from the volcano might hit the town of Shimabara at the foot of Unzen. But it was instead a pyroclastic flow — a very fast-moving mixture of gas and rocks — that engulfed unprepared researchers, journalists, firemen and policemen.

Daisuke Shimozuru, chairman of a government committee for predicting volcanic eruptions, defended his committee's record last week by saying that it had warned that pyroclastic flows might occur. The committee's public pronouncements, however, did not explain what a pyroclastic flow is, how deadly it is, or how fast it can strike.

Similarly, the Meteorological Agency, which is also responsible for issuing warnings, spoke repeatedly about the possibility of mudflows right up to the day of the disaster, but mentioned pyroclastic flows only in passing. And a few weeks before the eruption,

local government officials were quoted as saying that the lava flows from Unzen were slow moving and viscous — and therefore not so dangerous — and that their chief concern was mudflows.

This is not the first time a Japanese volcano has not behaved as predicted by government officials. In 1986, when the volcanic island of Oshima near Tokyo began rumbling, the volcanic prediction committee announced that an eruption was not imminent. A few days later, the eruption started. At that point, the committee predicted that lava flows would be confined to the volcano's caldera, and about 2,000 tourists flocked to the island to watch the spectacular eruptions. When the eruptions calmed down, the Meteorological Agency announced that activity seemed to be over.

A few hours later, a line of vents opened up across the caldera wall, lava poured down towards the main village, and 10,000 islanders (as well as the 2,000 tourists) had to be evacuated from the island overnight.

In 1989, when a series of earthquakes hit the resort city of Ito on the coast of Sagami Bay about 100 km from Tokyo, the government's earthquake prediction committee initially attributed the earthquakes to plate tectonic movement in the area. Local fishermen pointed out that brown water was bubbling to the surface of the sea and dead fish were being washed ashore. But this information was ignored. After earthquake activity calmed down briefly, the Meteorological Agency announced that the earthquakes seemed to be subsiding and warnings for residents were lifted.

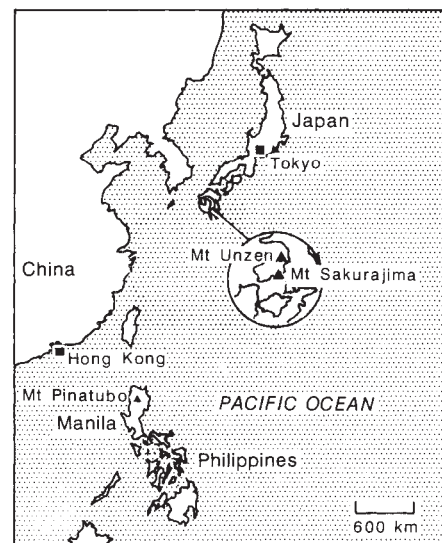
Hours later, strange booming sounds were heard emanating from the ground in Ito, and only then did the experts begin to talk seriously about the possibility that the earthquakes might be caused by magma rising to the surface. Before any predictions could be made, a submarine volcano broke through the sea's surface 1 km offshore.

The picture is further confused by the multitude of government-related organizations and committees — both national and local — involved in disaster prevention and prediction, a situation that tends to diffuse responsibility for public pronouncements. Perhaps the most surprising thing for Western observers is that, despite all that has happened in the past few years, the various disaster related organizations continue to make predictions and warnings, and the Japanese public continues to listen. **David Swinbanks**

■ Among those killed by last week's eruption of Mount Unzen were a French husband and wife team of volcanologists and a US geologist.

Maurice and Katia Krafft, famous for their daring pictures and documen-

Eruptions no secret



A series of volcanic eruptions in Japan and the Philippines in the past ten days may not have been the largest or most damaging volcanic events in history, but they were among the most heavily monitored. Over the past several years, researchers have installed instruments to watch for local tremors, gas emissions, swelling and thermal changes around the volcanoes. In the case of the Mount Unzen, which exploded earlier this month, measurements predicted an eruption two months before the event, which allowed officials to evacuate the area and minimize injuries.

The eruption last week of Mount Pinatubo in the Philippines and warning signs of a possible eruption at Mount Sakurajima in Japan were also predicted several weeks ago by volcanologists monitoring the volcanoes. Of the 550 active volcanoes worldwide, more than a hundred are currently monitored with some instrumentation, and several dozen have a full complement of earthquake, gas and size-change instruments to warn of impending eruptions, says US Geological Survey researcher Tom Cassadeval **C.A.**

ties of erupting volcanoes, flew to Japan just two weeks ago to film Mount Unzen because the erupting volcano provides a rare opportunity to observe pyroclastic flows in action.

They were accompanied by Harry Glicken, a US visiting postdoctoral fellow at Tokyo Metropolitan University, who formerly worked for the US Geological Survey and was an expert on Japan's volcanoes, in particular Mount Unzen.

All three were well aware of the dangers of approaching the volcano. Shortly before leaving for Unzen, Glicken said in Tokyo that the eruption was in many ways similar to that of Mont Pelée in Martinique in 1902, which destroyed St Pierre, the capital of the island and killed all 30,000 residents except for one prisoner in an underground cell. The three scientists, whose bodies were found last Thursday, had apparently ventured closer than anyone else to the foot of the erupting volcano. **D.S.**

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

A fireman and fire engine flee the flow of lava, rocks and ash at Mt Unzen.

Labour plans science revamp

London

THE British Labour Party promises a radical overhaul of the organization of science in the UK in "Pushing Back the Frontier", a policy statement on science and technology published Tuesday (11 June).

If the Labour Party is returned to power after 12 years in opposition (a general election must be called within a year), a Labour Government would appoint a Minister for Science in the Cabinet Office, responsible for the budgets of the five research councils, and co-ordinating their activities with other government departments.

The minister would be advised by a new Council on Science, Technology and Research, which will combine the present roles of the Advisory Council on Science and Technology (which now deals with general science policy issues across the whole of government) and the Advisory Board for the Research Councils (which now advises the Department of Education and Science on the research councils' budgets).

A Labour government would also restructure science in parliament, creating a new House of Commons select committee on Science, Technology and Research — the Commons committee that now oversees science also tackles education and the arts. This committee would be aided by a new Office of Technology Assessment, an independent body modelled on the US office of the same name.

Labour also proposes a package of fiscal incentives to encourage industry to invest in research and development (R&D). Following moves in the United States, Labour

would offer a 25 per cent tax credit for increases of R&D expenditure over a base year.

This credit would be over and above the current 100 per cent tax allowance for R&D.

These moves, says Labour science and technology spokesman Jeremy Bray, would help push civil R&D from the present 1.8 per cent of Gross Domestic Product toward a declared target of 2.5 per cent. Some defence R&D effort will be encouraged to move into the civil sector. Bray admits that tax incentives will not be enough to reinvigorate British science, but is reluctant to attach hard figures to real-term cash injections into science.

The Department of Education (stripped of its present responsibility for the research councils' budget) would be expected to work with the Higher Education Funding Council (a union of the Universities Funding Council and the Polytechnics and Colleges Funding Council, already preempted by the current Conservative government) on new ways to support basic science. The Labour Party also plans concessions for industry-sponsored R&D in universities.

An upturn in R&D would prompt increased demands on training, and the document proposes to establish an Advanced Certificate of Education and Training. The certificate would be a secondary-level qualification that would replace a "jungle" of academic and vocational qualifications (including "A"-levels) and encourage more students, particularly women, to take up science-based higher education, the document says.

Henry Gee

EUROPEAN SCIENCE FUNDING

Auditors decry EC research waste

London

THE European Communities' (EC) research programmes are wasting thousands of millions of pounds, according to a new report from the Court of Auditors, the body that oversees all European Commission spending. Only a tiny fraction of the 5,400 million ECU (about £3,800 million) budget spent in the past five years has resulted in concrete applications, it says.

The majority of EC-funded research is so-called "pre-competitive" industrial research. The idea is that companies will come together to work on underlying technology for their industry, and then work separately to turn this research into commercial products. But the Court found that few patents have arisen out of EC programmes.

The Court suggests that a greater proportion of the budget be devoted to exploiting and disseminating research results to a wider audience. Only 0.7 per cent (55 million ECU) of the budget of the second four-year Framework programme,

now coming to a close, was so earmarked.

Some commentators feel that the Court has been unnecessarily harsh in its judgments. British Labour Member of the European Parliament Gordon Adam, vice-chairman of the Parliament's Energy, Research and Technology Committee, agrees that the Court of Auditors fulfills a useful function. "We need to have a greater understanding of what we're achieving", he says, given the huge sums involved. But it is in the nature of many of the projects funded by the EC that they take some considerable time to come to maturity, and the Court may wish to reassess its opinions in one or two years' time. Pre-competitive research, says Adam, is "by definition" difficult to assess.

EC officials have also preempted the Court's criticisms on the proportion of funds devoted to technology transfer and the dissemination of research results: the funds allocated to this end have risen to 1 per cent in the third Framework Programme.

Henry Gee

UK military favoured

London

CONTINUING concern that the benefits of British spending on defence research and development are not effectively transferred to civil industry is underlined by a report from the Parliamentary Office of Science and Technology, published last week.

The office is the British equivalent of the US Congressional Office of Technology Assessment, but has a budget only about 1 per cent as big. The defence study is the second of the more substantial reports with which the office supplements its service of informal commentaries on current issues to members of the British Parliament.

The transfer of defence technology to civil industry has been a live issue in Britain since the end of the Second World War, partly because defence research and development is a large proportion of total government spending on research — it has fallen to 44 per cent of the total only in the past few years — and also a large proportion (now 20 per cent) of total spending on research by government and industry.

The parliamentary office injects a novel anxiety into the debate with the question whether the "inevitable" reduction of defence spending now in prospect will leave stranded many military research and development projects whose technology might be of value to civil industry. It is also concerned that the transformation of the defence research establishments into a government agency (from 1 April), which means that they have an incentive to reduce costs not directly related to contracts they win, may diminish their interest in technology transfer to civil industry.

While acknowledging that the present arrangement under which the Ministry of Defence is responsible exclusively for war and weapons has the virtue of avoiding muddled lines of responsibility, the office report says that the neglect of civil applications by the military risks "weakening further" the British technology base.

The most curious phenomenon brought to light is the fate of the organization called Defence Technology Enterprises, set up in 1985 with venture capital, and which appeared to be functioning well until 1990, when it "effectively ceased to operate". Tolerantly, the report notes that there were at the outset several reasons for scepticism about the enterprise, counting the gap between defence research and commercial application as the chief of these.

Wistfully, the office also noted that the British Technology Group might be well placed to take over the responsibilities of Defence Technology Enterprises were it not for the uncertainty created by the government's plans to privatise the group.

John Maddox

State of British science

SIR — Terence Kealey¹ argues that British science is growing, relative to other major industrialized countries, because cuts in government funding have been offset by increased support from industry and charities. This may be correct in Kealey's own field (clinical biochemistry), but it is certainly not true in general.

The United Kingdom is the only developed country in which government spending on civil science has decreased in recent years as a proportion of gross domestic product (GDP). In 1978–79, spending was 0.6 per cent of GDP; ten years later, in 1988–89, it was 0.52 per cent². A recent study³ shows that, in 1987, total government spending on civil research and development was £2,400 million in the United Kingdom, £4,670 million in France and £6,430 million in West Germany. As the populations of these countries are very similar (57 million, 56 million and 61 million respectively), there is a massive shortfall in per capita spending by the UK government relative to comparable European countries. Kealey claims that this shortfall is made up by industry and charities. In fact, when industrial funding of research and development is taken into account, the gap widens. (Industrial funding in 1987 in the United Kingdom was £4,700 million, in France £5,200 million and in West Germany £11,570 million. These figures are for industrial funding of both civil and defence research and development³.) Outside medical research, funding by charities is negligible.

Because of cuts in support for UK science in recent years, it is particularly difficult to fund important new developments. For example, most of the major developments in high- T_c superconductivity are from Japan and the United States. This is hardly surprising in view of the funding figures for superconductivity research. In 1989, for example, the United States (industry and government funding) spent £175 million, Japan £125 million, India £10 million and the United Kingdom £10 million⁴. The United Kingdom is therefore now funding an important research area at the same level as India. In 1989, the United Kingdom spent only 1.5 per cent of world expenditure on superconductivity research, and contributed about 1.5 per cent of the 15,000 publications on superconductivity. Thus, in science, as in other areas, you largely get what you pay for.

The low UK expenditure on superconductivity research is echoed in other areas of research. In the same issue of *Nature* in which Kealey's letter appeared, there was a news item⁵ in which it was reported that the United States planned a massive increase in the \$1,600 million that the federal government now spends on materials research. A number of states match such funding in state universities, so the total public funding of materials research in the United States is

probably about \$2,000 million. In the United Kingdom, by comparison, the total public funding last year of materials research was about £60 million⁶, that is, about 20 times lower than that in the United States. And support by the Science and Engineering Research Council of new research grants in materials and other areas is being halved this year relative to last year because of inadequate government funding.

All governments have a duty to the people they govern to protect the future of their country. A developed country is unlikely to survive economically without a strong science-based industry, and unless it plants the seedcorn of research it is unlikely to reap the harvest of production and profits. Science and technology are now advancing so rapidly, and the learning curve is so steep, that once a nation has fallen behind in a particular area, it is very difficult to catch up. A government that fails to invest adequately in scientific research is undermining the future of its country in a serious and fundamental way, and the damage is likely to be permanent and irreversible.

COLIN HUMPHREYS

University of Cambridge,
Department of Materials Science and
Metallurgy,
Pembroke Street,
Cambridge CB2 3QZ, UK

1. Kealey, T. *Nature* **350**, 370 (1991).
2. *Annual Review of Government Funded Research and Development 1990* (HMSO, London, 1990).
3. Atkinson, H., Rogers, P. & Bond, R. *Research in the United Kingdom, France and West Germany: A Comparison* (SERC, Swindon, 1990).
4. *Review of the UK National Superconductivity Programme* (Joint DTI/SERC report, in the press).
5. Anderson, C. *Nature* **350**, 365 (1991).
6. Humphreys, C. J. *Angew. Chem. Adv. Mater.*, **101**, 1103–1104 (1989).

SIR — Terence Kealey's letter is seriously misleading.

According to the Institute for Scientific Information¹, British performance as measured by the Citation Impact rating fell by 3.4 per cent over the period 1981–85 to 1986–90 while those of the United States, West Germany (as was), France and Japan have all risen.

Between 1980–81 and 1988–90, although there was a 7.5 per cent fall in the number of wholly university funded staff in science and engineering, a net increase of 13 per cent occurred because of a rise in short-term appointments. But about two-thirds of such posts are funded by research councils, not by industry or charities, and the majority are PhD students taking research assistant posts because postgraduate training grants are too low².

Funding by British industry and charities did increase between 1983 and 1988³ from about 10 per cent to 14 per cent of the total for research and development performed in higher education institutions (HEIs). But the

funds from charities are almost entirely for medicine, and the main industrial contributors are from the pharmaceutical sector; basic research, and especially the physical sciences, enjoy little benefit. The proportion of research at HEIs funded by industry has reached 6 per cent, about the same level as in Germany and the United States; it is unlikely to grow substantially greater.

The support for research in British HEIs as a fraction of GDP, or per capita (from all sources) is, according to the latest figures from the Organization for Economic Cooperation and Development, the lowest of nine European countries (including Switzerland). In West Germany it is more than 25 per cent higher — worth about £400 million a year.

Mrs Margaret Thatcher has explained⁴ why government has the major responsibility for funding the science base, and stressed the importance of 'curiosity-led' research. The dangers of neglect are taught by history — although not in the bizarre reading due to Kealey.

In Margaret Gowing's words⁵: "The doctrine of *laissez-faire* reached its zenith in Britain just as industrialisation accelerated into so-called revolution . . . Britain . . . had been very obviously outclassed in the Paris international exhibition of 1871, most notably by Germany . . . (where the) states had endowed science . . . by financing a system of universities . . . Yet despite increasingly urgent warnings the British government, averse from increasing public expenditure, did almost nothing." The consequences became apparent when the First World War broke out and Britain "found herself ignominiously dependent for crucial science-based products on imports from Germany". It was this rude awakening that led to the setting up of the Department of Scientific and Industrial Research, later to become the Science and Engineering Research Council.

The relentless decline⁶ as a fraction of GDP in the British government's funding of civil research, including the science base, while in other countries higher levels of government investment have been maintained, or reached, lies at the root of the present threat to the continued excellence of British science, and to its capacity to support a future high-technology based economy⁷.

JOHN H. MULVEY

Save British Science,
PO Box 241, Oxford OX1 3QQ, UK

1. *Science Watch* January 1991.
2. Evidence to the House of Commons Select Committee on Education, Science, and the Arts by M. Sharp, Save British Science, in the Committee's Report *Science and the European Dimension* (HMSO, London, Dec. 1990).
3. *Annual Reviews of Government Funded R&D* (HMSO, London, 1986–90).
4. Speeches to the Royal Society (Sept. 1988) and the Parliamentary and Scientific Committee (Dec. 1989).
5. 'An old and intimate relationship', Spencer Lecture 1982 (Oxford Univ. Press, published in *Science and Politics* 1984).
6. *British science: Benchmarks for the year 2000* (Save British Science, Oxford, 1990).
7. 'The economic value of basic science', Kay, J. A. & Llewellyn Smith, C. H., *Fiscal Studies* **6**, no. 3 (1986).

Opinions from an inquiry panel

The members of the panel at Tufts University that investigated the *Cell* article of which Dr Thereza Imanishi-Kari was the chief author respond to Dr Margot O'Toole's statement.

DOCTOR Margot O'Toole's response to the Office of Scientific Integrity (OSI), reprinted in part in *Nature* (351, 180; 1991), requires comment.

Early in May 1986, Dr O'Toole asked us to look into a potential controversy between herself and her postdoctoral adviser at the Massachusetts Institute of Technology, Dr Thereza Imanishi-Kari. We were not asked to, nor did we, investigate possible fraud. We considered that we were all friends and colleagues. Although we suggested that she pursue these issues with Dr Baltimore and other senior officials at MIT, she preferred at the time to have us resolve this scientific controversy so that she would not be directly involved.

Dr O'Toole made three claims: data in the laboratory records showed that a normal mouse expressed a high frequency of transgene-like idiotype; the Bet-1 reagent used was not as specific as claimed; and early sampling of transgenic mouse hybridoma supernatants led to an underestimation of the amount of transgene-encoded antibody. She felt that these problems raised doubts about the correctness of the conclusions of the *Cell* paper.

At a meeting with Dr Imanishi-Kari, we viewed data that resolved all three issues. A particularly troubling issue — the data Dr O'Toole brought to us indicating that a "normal" mouse produced a high frequency of transgene idiotype positive hybridomas — was dealt with conclusively by evidence that the mouse in question was a transgenic littermate.

We also reviewed data from radioimmunoassays of Table 2 clones. Although she now writes that she can "attest to the falsity of this claim", Dr O'Toole was not present at this meeting and cannot provide evidence in relation to it. The need to review these data arose from Dr O'Toole's concern that Table 2 hybridoma wells had been sampled too early. When we raised this issue with Dr Imanishi-Kari, she said that Table 2 data were derived from uncloned wells, but the hybridomas had been cloned and their antibodies characterized. She said many of the hybridomas produced idiotype positive antibodies lacking the transgene encoded allo-type. We asked if we could see the data. She asked if we didn't trust her. We were silent, and she began to cry as she brought out the data for us to examine. Our examination of the records convinced us that the necessary data existed and that Dr O'Toole's concerns could be answered. Contrary to the implication of her current statement, these two issues (and the Bet-1 questions) were

resolved to our satisfaction.

About a week after the meeting with Dr Imanishi-Kari, Dr O'Toole told us that she now believed the conclusions of the paper were completely wrong. We arranged a second meeting with Dr Imanishi-Kari.

Dr O'Toole currently claims that at the second meeting, also in May 1986, attended by Drs Imanishi-Kari, O'Toole, Huber and Wortis, "[Dr Imanishi-Kari] candidly admitted that certain crucial experiments described in the paper had not even been performed and that experiments that had been performed had not yielded the results claimed." This is simply not true. Dr Imanishi-Kari repeated a statement she had made previously, namely, that the hybridomas from Table 3 were isotypic; not those in Table 2 as written in the *Cell* paper (a correction that has been published). There were absolutely no statements made that hybridomas of Table 2 wells were never cloned and radioimmunoassays on the antibodies of such clones never performed.

Dr O'Toole did not raise with us the issue of the failure to clone and characterize hybridomas from Table 2 wells. Corroboration of our view of these events can be found in Dr O'Toole's memo to Dr Eisen, written within two weeks of our second meeting, as it does not refer to the cloning of Table 2 clones or the RIA analysis of their antibodies as issues. Thus there is no basis for confusion about the matters we dealt with and the current allegations of the OSI. Dr O'Toole's *Nature* statement, which suggests that the questions she raised to us in 1986 correspond to the crucial evidence in the current OSI draft report, is incorrect.

The focus of the second meeting with Dr Imanishi-Kari was data arising from transgenic mouse hybridomas that produced IgG and IgA antibodies which were idiotype positive. At the close of the meeting, Dr Imanishi-Kari said to Dr O'Toole (in these, or similar words), "if you want to believe that the data can be explained on the basis of hybrid formation between γ and μ chains, where the amount of μ message is undetectable on Northern and where you postulate an insensitivity of the Bet-1 (anti- μ) reagent, you are free to believe." Whereupon Dr O'Toole stood up and said, "I'm satisfied," and offered to shake hands with Dr Imanishi-Kari. There was absolutely no mention of a retraction.

Dr O'Toole's current statement in *Nature* restates her view. "Low-level production of the transgene could explain why the hybridomas were positive for the idiotype of the transgene."

We did not tell Dr O'Toole that the paper "would be retracted" nor did we agree that it should. We did not tell Dr O'Toole that "the protection of careers must take precedence over scientific accuracy." We never "slandered" or "libelled" Dr O'Toole.

BRIGITTE T. HUBER
Associate Professor
Tufts University School of Medicine
ROBERT T. WOODLAND
Associate Professor
University of Massachusetts
School of Medicine

HENRY H. WORTIS
Professor
Tufts University School of Medicine

I was Margot O'Toole's thesis adviser and maintained a sustained friendship with her. I had every reason to take her concerns seriously and to be supportive of her efforts to resolve scientific questions. In the end I disagreed with her assessment of the science. Different opinions are neither slanders nor libels, and I find it especially painful to hear my opinions so characterized by my first graduate student.

HENRY H. WORTIS

Dr O'Toole approached me with her disturbing findings because she trusted me as a friend and respected me as a scientist. I had served on her thesis committee and had maintained contact with her during her postdoctoral years. I considered her allegations seriously and reviewed her charges to the best of my abilities. I could detect no evidence of fraud in the material I reviewed. The fact that we disagreed in our interpretation of the data is no justification for impugning my honesty in evaluating it. I acted conscientiously, as a responsible scientist, with a commitment to science, and above all, out of friendship towards Dr O'Toole. In return, she has misrepresented my statements and deeds to the point where I now feel that she has defamed my scientific and personal reputation.

BRIGITTE T. HUBER

Although not personally accused by Dr O'Toole of slander and libel, I find offensive her statement that I falsely reported what I had seen during the data analysis. It is telling that the one constant to emerge from this unhappy episode is that anyone who has disagreed with Dr O'Toole's analysis of the scientific issues has been accused of either incompetence or deceit. Unfortunately, it is unlikely that the acrimony in this case will be soon reduced to a level where the facts can emerge.

ROBERT T. WOODLAND

Verification of the chemical convention

S. Black, B. Morel and P. Zapf

A chemical convention banning the production of chemical warfare agents is needed to control the activities of the chemical industry. But compliance must be in the industry's own interest.

AFTER many years of negotiations, an international convention banning the production, possession and use of chemical warfare agents (CWAs) is close to completion. To be a success, the chemical weapons convention will have to possess a verification system intrusive and powerful enough to ensure that in the vast world of chemical industry no violations take place. According to the FBI, hundreds of millions of dollars change hands today because of industrial espionage. The planned system of on-site inspections, which consists of granting inspectors essentially unlimited access to all information and activities in a plant, would exacerbate this problem.

Because on-site inspections will be an unavoidable part of the verification procedure, the problem is to find a system of inspections, at once adequate for the purpose of the convention and not damaging to the industrial sector. Here we propose one possibility. Inspections would be carried out first on waste streams¹, and only if suspicious evidence is found would the inspection be allowed to proceed.

Inspection of plant wastes is feasible because of the sensitivity of analytical chemistry techniques. A gas chromatograph coupled to a mass spectrometer can be used as a mobile inspection system sensitive to one nanogram or less of a given compound^{2,3}, which in a sample of one kilogram of wastes, corresponds to a mass concentration of 10⁻¹².

The use of gas chromatograph (which separates molecules on the basis of their retention time) and of a mass spectrometer (which recognizes molecules by measuring the mass and relative abundance of their ionization products), requires an exact list of target molecules. Most of the molecules in the waste stream are hydrolysed or undergoing chemical reactions. The use of wastes as a source of information supposes that the degradation properties of the compounds of interest are known.

Waste production

Wastes are produced after each separation stage of manufacturing. Separation levels (for example, in distillation columns), such that not even one nanogram of the compound purified could be found in every kilogram of waste generated (in that case distillate⁴), are not always possible. They require that the compounds to be separated have suitable physical properties (very different volatilities). They also require expensive equipment, which to achieve such a degree of separation would have to be operated for a

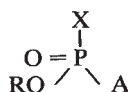
very long time, making the manufacture extravagantly expensive.

The detection in the wastes of CWAs would constitute a clear signal of noncompliance. But in many interesting cases (three out of four nerve agents), the CWA can be a binary weapon. What leaves the plant is a relatively harmless precursor which gives rise to a nerve agent when mixed with another ingredient in the warhead, just before the CWA is released.

In other cases, the precursors are so common (mustard, for example, is made by mixing thiodiglycol with hydrochloric acid), that they could meet accidentally in the wastes, and produce some CWAs, spuriously. Furthermore, the production of mustard can literally be made in a saucepan. In this case, even full-scale on-site inspections would suffer from a high probability of false negatives. But mustards need to be used in much larger quantities than nerve agents to be militarily effective. It makes sense to rely on on-site inspections for those CWAs like nerve agents whose manufacture is not trivial or which have a well-defined chemistry. And the convention would gain in cost effectiveness by using a different methodology of verification for different compounds instead of trying to treat all CWAs in the same way when there is no verification technique which applies equally well to each.

Organophosphorus nerve agents are the most toxic (by inhalation and by skin penetration) of the existing CWAs. They fall into three classes: tabun, sarin/soman and the V-agents (VX). Their toxicity measured as the test lethal dose to mice with 50 per cent probability (LD₅₀) varies from 0.27 mg per kg for tabun to 0.01 mg per kg for VX. (These compounds inhibit the hydrolysis of acetylcholine by acting on one or both of the active sites of the enzyme acetylcholinesterase⁶.)

The chemical structure of the organophosphorus nerve agents has the general form



With R = C₂H₅, A = CN, X = (CH₃)₂N for tabun; R = isopropyl, A = methyl, X = F for sarin; R = isobutyl, A = methyl, X = F for soman; R = ethyl, A = methyl, X = SCH₂CH₂N (i-C₃H₇)₂ for VX.

Tabun was the first nerve agent created. Sarin and soman differ only by the ester (isopropyl for sarin and isobutyl for soman), which gives them different volatilities. Sarin,

soman and VX can be produced as binary weapons, whereas tabun cannot.

Organophosphorus nerve agents are conspicuous by their P-C bonds. Although a P-C bond is not more energetic than the other phosphorus bonds (the P-C bond is around 70 kcal mol⁻¹, the P-O bond is roughly the same, the P=O bond is 120 kcal mol⁻¹, the P-F bond is around 110 kcal mol⁻¹ (ref. 7), the P-C bond tends to have a much higher activation energy, especially when the alkyl is methyl (this does not apply to tabun).

Difficult bonds

As a result, in all the cases except for tabun, the creation of the P-C bond (alkylation) tends to be the most difficult phase of the production of nerve agents: alkylation methods used for the production of VX include burning methane at 600 °C in phosphorus trichloride (PCl₃), with a yield of only 15 per cent. The P-methyl bond tends also to be the most persistent remnant of the molecule after degradation. Furthermore, although there are many molecules with a P-C bond⁸, few commercial products have a P-methyl bond. Commercial organophosphorus compounds tend to have esters instead of alkyl groups. Most organophosphorus compounds tend to be produced from phosphorus trichloride (PCl₃), trimethyl phosphite (P(CH₃O)₃), or phosphorus oxichloride (POCl₃).

Both sarin and soman are made using the same reactions. A different alcohol needs to be substituted as the ester group for the production of soman (isobutanol instead of isopropanol for sarin). One of the several production processes for sarin is as follows:

- i $\text{PCl}_3 + 3\text{CH}_3\text{OH} \rightarrow (\text{CH}_3\text{O})_2\text{P}(\text{OH}) + 2\text{HCl} + \text{CH}_3\text{Cl}$
- ii $(\text{CH}_3\text{O})_2\text{P}(\text{OH}) \rightarrow (\text{CH}_3\text{O})\text{P}(\text{O})(\text{CH}_3)(\text{OH})$
 $2(\text{CH}_3\text{O})\text{P}(\text{O})(\text{CH}_3)(\text{OH}) \rightarrow (\text{OH})(\text{CH}_3)(\text{O})\text{POP}(\text{O})(\text{CH}_3)(\text{OH})$
- iii $(\text{CH}_3\text{O})\text{P}(\text{O})(\text{CH}_3)(\text{OH}) + 2\text{PCl}_3 + 2\text{Cl}_2 \rightarrow$
 $\text{CH}_3\text{P}(\text{O})\text{Cl}_2 + \text{CH}_3\text{Cl} + \text{HCl} + \text{POCl}_3$
 $(\text{OH})(\text{CH}_3)(\text{O})\text{POP}(\text{O})(\text{CH}_3)(\text{OH}) +$
 $3\text{PCl}_3 + 3\text{Cl}_2 \rightarrow 2\text{CH}_3\text{P}(\text{O})\text{Cl}_2 + 2\text{HCl} + 3\text{POCl}_3$
- iv $2\text{CH}_3\text{P}(\text{O})\text{Cl}_2 + 2\text{HF} \rightarrow \text{CH}_3\text{P}(\text{O})\text{Cl}_2 +$
 $\text{CH}_3\text{P}(\text{O})\text{F}_2 + 2\text{HCl}$
- v $\text{CH}_3\text{P}(\text{O})\text{Cl}_2 + \text{CH}_3\text{P}(\text{O})\text{F}_2 + 2\text{i-C}_3\text{H}_7\text{OH} \rightarrow$
 $2(\text{i-C}_3\text{H}_7\text{O})\text{P}(\text{O})(\text{CH}_3)\text{F} + 2\text{HCl}$

The final step is done without purifying the didi (dichloro/difluor) mixture from iv and can take place in a warhead. The HCl occurs as a gaseous byproduct, and no solvents are required for this reaction. The yield of this process (iv + v) is about 87 per cent (ref. 9).

The Newport process, the only large-scale process used to produce VX, is as follows:

- i $\text{PCl}_3 + \text{CH}_3\text{H} \rightarrow \text{CH}_3\text{PCl}_2 + \text{HCl}$
- ii $\text{CH}_3\text{PCl}_2 + 2\text{C}_2\text{H}_5\text{OH} \rightarrow \text{CH}_3\text{P}(\text{OC}_2\text{H}_5)_2$
- iii $\text{CH}_3\text{P}(\text{OC}_2\text{H}_5)_2 + \text{HOCH}_2\text{CH}_2\text{N}(\text{i-C}_3\text{H}_7)_2 \rightarrow \text{C}_2\text{H}_5\text{O.CH}_3\text{P.OCH}_2\text{CH}_2\text{N}(\text{i-C}_3\text{H}_7)_2 + \text{C}_2\text{H}_5\text{OH}$
- iv $\text{C}_2\text{H}_5\text{O.CH}_3\text{P.OCH}_2\text{CH}_2\text{N}(\text{i-C}_3\text{H}_7)_2 + \text{S} \rightarrow \text{C}_2\text{H}_5\text{O.CH}_3\text{P(S).OCH}_2\text{CH}_2\text{N}(\text{i-C}_3\text{H}_7)_2$
- v $\text{C}_2\text{H}_5\text{O.CH}_3\text{P(S).OCH}_2\text{CH}_2\text{N}(\text{i-C}_3\text{H}_7)_2 \rightarrow \text{C}_2\text{H}_5\text{O.CH}_3\text{P(O).SCH}_2\text{CH}_2\text{N}(\text{i-C}_3\text{H}_7)_2$

The heat of the sulphidation reaction iv is enough to drive the thiono–thiolo rearrangement v. In a binary configuration, iv and v take place in the warhead.

To detect the manufacture of sarin, soman or VX, it is necessary to be able to detect the manufacture of their precursors as binary weapons. The signal to be detected should reveal the production of didi for sarin and soman (end of stage iv), and of the methyl compound produced at stage iii of the production of VX (referred to as QL). One should exploit the fact that the P–methyl bond is very rare in the industrial world and very resilient to chemical treatment. The other bonds are much less resilient. The molecules to be detected are the nerve agents themselves, the precursors we just mentioned and their degradation products in the waste stream. The P–methyl bond will appear in the degradation products of the compounds of interest in the form of methylphosphonic acid and other derivatives.

Tabun is a special case in that alkylation is, in general, the last stage of the production process. Furthermore, the P–C bond in this case is not particularly strong (P–CN does not have a very large activation energy, as it degrades easily). But tabun cannot be made into a binary weapon, and it can be found in the waste stream. The production of tabun from dimethylamine is as follows (this is the process used by the Germans during the Second World War, and reportedly the process used by the Iraqis):

- i $(\text{CH}_3)_2\text{NH} + \text{POCl}_3 \rightarrow (\text{CH}_3)_2\text{NP(O)Cl}_2 + 2\text{HCl}$
- ii $(\text{CH}_3)_2\text{NP(O)Cl}_2 + \text{NaCN} + \text{C}_2\text{H}_5\text{OH} \rightarrow (\text{CH}_3)_2\text{NP(O)(C}_2\text{H}_5\text{O)(CN)} + \text{HCl} + \text{NaCl}$

To prevent thickening of the reaction product due to the formation of salt (in step i), POCl_3 is used in excess and a 60 per cent $(\text{CH}_3)_2\text{NP(O)Cl}_2$ solution is obtained. Fractional distillation produces a 90 per cent yield of 99.5 per cent pure $(\text{CH}_3)_2\text{NP(O)Cl}_2$ (ref. 10). The reaction product of Germany's production of tabun in the Second World War (step ii) is reported to have contained 25 per cent tabun together with 75 per cent chlorobenzol. An 83 per cent yield of tabun is obtained by vacuum distillation. The final product was originally stored 95 per cent pure (the rest chlorobenzol), and then 80 per cent pure for filling into munitions. The 80 per cent pure solution with 20 per cent chlorobenzol was found to be more stable¹⁰.

A second production process for tabun,

developed by the British at the end of the Second World War, is:

- i $(\text{C}_2\text{H}_5\text{O})_2\text{PCl} + 2\text{CH}_3\text{NH} \rightarrow (\text{C}_2\text{H}_5)_2\text{PN}(\text{CH}_3)_2 + (\text{CH}_3)_2\text{NH} + \text{HCl}$
- ii $(\text{C}_2\text{H}_5\text{O})_2\text{PN}(\text{CH}_3)_2 + \text{ICN} \rightarrow (\text{CH}_3)_2\text{NP(O)(C}_2\text{H}_5\text{O)(CN)} + \text{C}_2\text{H}_5\text{I}$

Both of these steps require separation of tabun from either unreacted precursors or a solvent. Only the waste stream of the final step will contain a P–C bond, in the form of tabun itself, which therefore would have to be used as the signal.

The degradation by hydrolysis of tabun is rather fast and the CN group cleaves off the molecule first^{11,12}. After 20 hours, Finnish tests showed only 2.5 per cent of the original amount of tabun remaining (lake water, pH 5.5, room temperature). Other sources show 99.9 per cent decomposition time in distilled water (25 °C) of 14–48 hours¹².

The primary degradation product of tabun is ethyl hydrogen *N,N*-dimethylphosphoramide. This molecule's existence would suggest tabun production. Hydrolysis of this product is relatively slow. The secondary degradation products $(\text{CH}_3)_2\text{NP(O)(OH)}_2$ and $\text{C}_2\text{H}_5\text{OP(O)(OH)}_2$ are not as useful for analysis, as they occur as degradation products from common chemicals¹¹.

To escape detection of the manufacture of nerve agents, a level of concealment in the wastes of all indications is required that would be totally unachievable, especially in industrial conditions. Chemical treatments or incineration even in combination would not be able to erase all traces of circumvention at any reasonable cost (the Johnston Island incinerator which will be used to destroy existing CWA stocks will destroy with an efficiency of 10^{-7} ; ref. 13).

Thus an inspection failing to detect any signal of nerve agent manufacture in the waste stream could be terminated without any loss of verifiability for the convention with respect to the present system. It would be only when a signal is detected, that further investigation would be necessary.

But positive signal does not necessarily imply noncompliance. False positives could occur in the case where the plant is manufacturing or using a compound involving a P–methyl bond. These cases are rare in the industrial world. The principal compound we found is dimethyl methylphosphonate (DMMP), produced by the molecular rearrangement of trimethyl phosphite¹⁴. DMMP is a flame retardant, a precursor for other flame retardants, and could be used as an intermediate in the production of nerve agents. The production of DMMP is on the order of 2,200 tonnes worldwide among four producers¹⁵, including Akzo and Tenneco Inc.¹⁶. Fonotos is one of the very rare organophosphorus pesticides, with a p-alkyl bond and the alkyl is not methyl, but ethyl (the formula is $(\text{C}_2\text{H}_5\text{O})\text{C}_2\text{H}_5\text{P(S)S(phenol)}$). One of the precursors in the production of Fonotos is dichlor, the closest chemical to a nerve agent that is manufactured commercially.

There may be a few more examples, but these would not affect the conclusion that waste-stream analysis would reduce dramatically the number of plants which could have to undergo a full-scale inspection. The detection of methylphosphonic acid in the wastes should not necessarily trigger a full-scale investigation. The absence of fluorine in the waste (fluorine is very conspicuous) would exclude the possibility that sarin or soman were being manufactured. VX would then become the only possible nerve agent being produced. VX does not hydrolyse as fast as most other nerve agents, and if manufactured there it (or its last precursor QL) would presumably be present in the wastes as well. It is up to the plants to anticipate that their waste stream will contain this kind of signal and to design a strategy, when possible (and that is in general the case¹⁷), to provide the inspectors with compelling evidence that nothing illegal is taking place, without the need of a full-scale investigation. Obviously there might be situations where full investigations would be required. But, instead of being the rule, they would be the exceptions. The impact of that kind of inspection on the industry would be much more limited than the proposed system. As a result the convention could be more verifiable, because it could hope to benefit from the cooperation of industry. □

S. Black, B. Morel and P. Zapf are in the Department of Engineering and Public Policy, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA.

1. Furguson, T. L., Hylton, A. R. & Mumma, C. E. *Studies on the Technical Arms Control Aspects of Chemical and Biological Warfare* (Midwest Research Institute, Kansas City, Missouri, 1972).
2. Smith, R. *Gas and Liquid Chromatography in Analytical Chemistry* (Wiley, Chichester, 1985).
3. *Air Monitoring as Mean for Verification of Chemical Disarmament* (Ministry of Foreign Affairs of Finland, Helsinki, 1985).
4. Douglas, J. M. *Conceptual Design of Chemical Processes* (McGraw-Hill, New York, 1988).
5. *The Problem of Chemical and Biological Warfare* Vol. 5 (Stockholm International Peace Research Institute, Almqvist & Wiksell, Stockholm, 1971).
6. Benschop, H. P. & DeJong, L. P. *Acc. Chem. Res.* **21**, 368–374 (1988).
7. Sanderson, R. T. *Chemical Bonds and Bond Energy* (Academic, New York, 1971).
8. Rautio, M. in *Non-production by Industry of Chemical Warfare Agents: Technical Verification under a Chemical Weapons Convention* (ed. Lundin, S. J.) 77–89 (Stockholm International Peace Research Institute (SIPRI), Oxford University Press, Oxford 1988).
9. Robinson, J. P. (ed.) *The Chemical Industry and the Projected Chemical Weapons convention* (proceedings of a SIPRI/Pugwash conference, Oxford University Press, 1986).
10. Lohs, K. in *Synthetic Poisons (Chemistry, Effects and Military Significance)* (East German Military Publishing House, 1963).
11. *Identification of Degradation Products of Potential Organophosphorus Warfare Agents* (Ministry of Foreign Affairs of Finland, Helsinki, 1980).
12. *Review of Reactions of Chemical Agents in Water* (US Army Biomedical Research and Development Laboratory, AD-A213287, 1989).
13. Emle, L. *Chem. Engin. News* (13 August 1990).
14. *Encyclopedia of Chemical Technology* 3rd Edn. Vol. 10, (eds Kirk, R. E. & Othmer, D. F.) 400 (Wiley, New York).
15. Chemical Manufacturers Association (letter to P. Zapf, 24 October 1991).
16. *Directory of Chemical Producers* (SRI International, Menlo Park, California, 1989).
17. Black, S., Morel, B. & Zapf, P. *Technical Aspects of Verification of the Chemical Weapons Convention*. (Carnegie Mellon University, Pittsburgh, 1991).

Can reason defeat unreason?

This is Animal Rights Awareness Week in the United States – an occasion celebrated by the groups calling themselves animal activists of one stripe or another. Among biomedical scientists, the ‘celebration’ occasions acute despair.

DURING the past decade, US animal rights activists, initially dismissed by many researchers as on the fringe, have gained a strong political foothold, with an estimated 10 million individual members nationwide and a war chest of at least \$50 million. As Columbia University medical dean Herbert Pardes and colleagues wrote in a recent issue of the *New England Journal of Medicine*, “Conventional wisdom may see animal activists as the scrappy underdog grappling against the ‘biomedical behemoth backed by massive resources.’ Conventional wisdom is wrong.”

A small corps of researchers is trying to fight back, but recent efforts simply reveal what a long way they have to go. In an attempt to build a broad congressional base for their cause, a coalition representing more than 50 health organizations gave a party at the US Senate last week. Called “Saving Lives”, the reception drew a couple of hundred people from member organizations and one senator. He, Senator Howell Heflin (Democrat, Alabama) drew applause by announcing that he will reintroduce a bill to make it a federal crime to break into a research laboratory or make unauthorized use of research records, data or animals.

The animal rights groups can be expected to deluge Congress with letters of opposition from millions of ordinary citizens who have been convinced that research scientists are cruel. It is inconceivable that researchers could drum up letters of support that would even be worth counting. And therein lies part of the problem. The other lies in the scientists’ overall approach.

Timed to catch Congress’s attention in advance of lobbying by activists during Animal Rights Awareness Week, the biologists were scooped even in this effort by People for the Ethical Treatment of Animals (PETA), which staged an animal-free circus at the House of Representatives the night before.

Where PETA had people painted as tigers handing out vegetarian tofutti bars, the coalition for Saving Lives had speeches. Surgeon General Antonia Novello, a paediatrician, talked about the animal studies that made the polio vaccine possible, and noted continuing research for a vaccine to save 1,000 infants a year aged 18 to 24 months from death from influenza type B. “All of this is due to animal research”, she said sincerely. The audience nodded politely. Nobel laureate Harold Varmus spoke compellingly about the role of animals in cancer research. Celebrity heart surgeon Michael DeBakey made the pitch for cardiology. Collectively

they made a strong case, but they were preaching to the converted.

The animal rights people go for the heart, the biologists for the head. Just recently, the National Academy of Sciences (NAS) and the Institute of Medicine (IOM) released a highly reasoned glossy ‘white paper’ (policy document) called *Science, Medicine, and Animals* that, inexplicably, has a photomicrograph of the structure of AZT on the cover rather than an animal whose sacrifice for research saved some child from imminent death. The white paper is an exemplary catalogue of the contributions of animal research to human medicine. It even makes a stab at touching the heart strings with brief vignettes. Gregg Mass, we are told, survived a case of non-Hodgkins lymphoma diagnosed just two weeks before his thirtieth birthday because of animals that died for him during years of research on chemotherapy.

But, in the end, the NAS/IOM monograph is meant for the head. “In writing this position paper”, say NAS president Frank Press and IOM president Samuel Thier, “our intention has been not to end the debate on whether and how animals are used in research; rather it has been to inform that debate.”

That is precisely the deficiency of the scientists’ position. They think they are engaged in an intellectual conversation. Not so. The past decade is evidence enough that the animal rights movement is not about reason. It is about eliminating the use of animals in research.

In 1985, the United States enacted the Health Research Extension Act which requires peer reviewers to ask whether the proposed use of animals in a protocol is necessary. That same year, the Animal Welfare Act of 1966 was amended to set standards for animal research and to establish for animals the equivalent of the institutional review boards that now exist at all research institutions and hospitals for the approval of human experimentation. Millions of dollars have been spent to improve facilities and correct deficiencies in research animal care. But it is not enough, largely because, for the extremists who, in the end, are in control, it is not the point.

Speaking at the Saving Lives reception, television actor David Birney got it right when he told the assembled researchers, “You have a powerful enemy abroad in the land that wishes you ill.” Scientists talk about facts. The animal people present their pitch through a photograph of a beautiful actress with a puppy in her arms. Sex and innocence.

Birney called it “deviously sentimental, manipulative, shabby, shameless.” By implication he acknowledged that it is effective.

It is often said in medicine, particularly with regard to addictive diseases, that the first step to a solution is an accurate diagnosis and recognition of the problem. For a couple of years now, research psychiatrist Frederick K. Goodwin, director of the Alcohol, Drug Abuse, and Mental Health Administration, has been at the forefront of the “tell it like it is” school of combat in the animal rights war. He bluntly accuses PETA and other extreme groups of engaging in disinformation campaigns and misleading emotionalism, and charges them with obfuscating their true goal of eliminating animal research altogether under the guise of doing only animal research that is absolutely necessary.

Goodwin, nearly alone among federal officials until recently, directly challenges the animal rightists’ underlying belief that animals and human beings are morally equivalent. Then, several months ago, Health and Human Services Secretary Louis Sullivan entered the fray by calling certain animal groups “terrorists”.

Now, Bernadine Healy, barely two months into her new job as director of the National Institutes of Health, has joined the crusade. Speaking at Saving Lives, Healy, who as a cardiologist believes that in many circumstances rats are more valuable for research than dogs, subtly raised the rhetoric coming from government when she said “animal activists espouse a fallacy, namely that medical progress can be maintained without essential animal research.” Healy decried the activists’ terrorism (she said death threats against researchers are “inhuman”) and worried about the success of rights groups in getting their films and magazines into public libraries and schools. All worth worrying about.

Yet within the research world there is still no well-honed campaign to fight fire with fire. The activists show pictures of animals in surgery and the surgeons look cruel. No one shows pictures of children in surgery (which is often bloodier by far), who then emerge alive and well. One is left to imagine that the pretty child who has had a liver transplant never had her abdomen cut open in the process. The activists do well by filming the before but not the after.

Meanwhile, the research community at large has yet to learn that some arguments cannot be won by dry, safe, reasoned discourse.

Barbara J. Culliton

Parent structure superconducts

Robert J. Cava

As anyone in the high-temperature-superconductivity game will tell you, in the past year or two it has become more and more difficult to find new copper-oxide-based superconductors. The days of the random lucky shot seem to be over: all the combinations of two or three non-transition element oxides with copper oxide have probably been tried ten times over. This is what makes the discovery of superconductivity below 40 K in $\text{Sr}_{1-y}\text{Nd}_y\text{CuO}_2$, reported by Smith *et al.* on page 549 of this issue¹, so exciting. No lucky shot, this. Rather the authors succeeded through patient detective work and careful chemical reasoning.

The chemical complexity involved in finding truly new materials with five or more oxide components, not forgetting oxygen stoichiometry, is intimidating to even the most seasoned crystal chemist. The problem, of course, is that in complex materials nature is working to balance subtle many-body interactions by a set of rules which we understand in only the simplest of cases. It is considered a theoretical triumph to be able to explain the crystal structures of the known binary compounds from first principles; the prediction of the structures of new compounds made from more than a few elements is simply out of the question. Certainly, by making more and more reliable measurements on better and better samples of 'old' superconductors, physicists are chipping away at the mysteries of the superconducting copper oxides. But virtually all leaps in understanding have followed the discovery of new materials, the recent paucity of which is having a predictable effect.

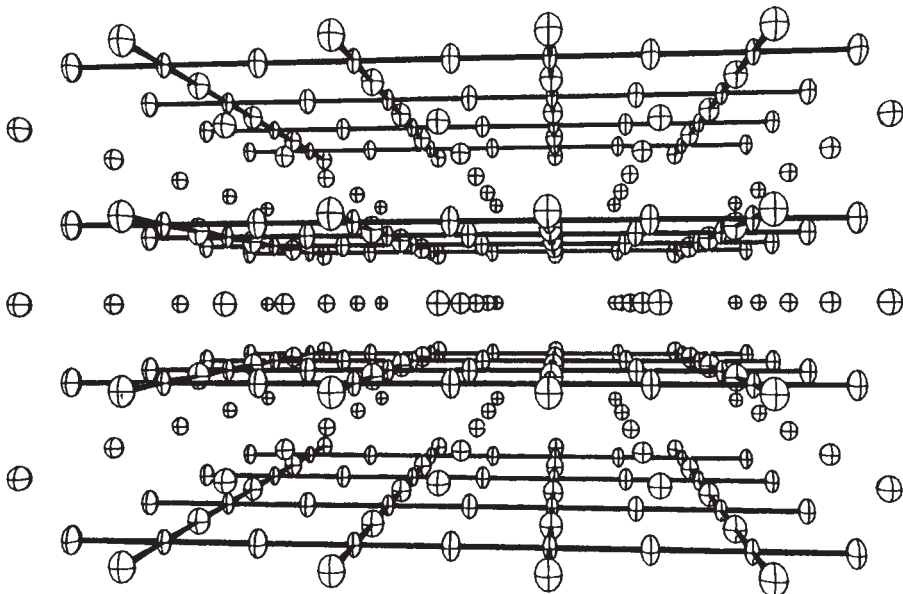
The work by Smith *et al.*¹, on electron-doped superconductivity in the 'all layer' compound $\text{Sr}_{1-y}\text{Nd}_y\text{CuO}_2$, is a good example of the kind of materials-science research that seems to be necessary to find new copper-oxide-based superconductors these days. It brings to fruition the authors' idea that the previously well-known 'parent' phase, based solely on copper-oxide planes, has all the right characteristics to be an electron-doped superconductor if the appropriate chemical and synthetic conditions could be found. Many groups have tried, but failed, to find the right recipe. Smith *et al.* have succeeded by employing the 'geological' synthetic conditions of 1,000°C and an applied pressure of 25 kilobars. Although the apparatus for doing such synthesis was commonly used during the heyday of solid-state oxide chemistry 20 years ago, it is now relatively rare, with at most two or three groups using it to investigate copper-oxide-based superconductors.

The development and successful execution of the idea is particularly interesting. It was clear, relatively soon after the high-temperature superconductors were identi-

fied, that the CuO_2 planes are the electronically active heart of superconductivity in copper oxides. For the 'hole-doped' superconductors, in which the formal copper valence is greater than two, there are apical oxygen atoms above and also sometimes below the copper oxide planes, making the copper five or six coordinated to oxygen. $\text{Nd}_{2-y}\text{Ce}_y\text{CuO}_4$, the first electron-doped (formal copper valence less than two) superconductor², was one of the new materials whose

Starting from their understanding of the structure of $\text{Ca}_{0.84}\text{Sr}_{0.16}\text{CuO}_2$, and the compressibilities of alkaline earth and CuO_2 arrays, Takano and coworkers⁴ discovered in 1989 that the all-layer phase could be stabilized for large alkaline earths by synthesis under high applied pressures. They found that SrCuO_2 and even $(\text{Ba},\text{Sr})\text{CuO}_2$ solid solutions could be synthesized. Owing to the larger sizes of these alkaline earths, the lengths of the Cu-O bonds in the new materials were significantly larger than those in $\text{Ca}_{0.84}\text{Sr}_{0.16}\text{CuO}_2$.

Smith *et al.* realized that the stretched CuO_2 array in SrCuO_2 is similar in size to that in $\text{Nd}_{2-y}\text{Ce}_y\text{CuO}_4$, and therefore more likely



The parent structure of all copper oxide superconductors³, showing infinite CuO_2 square-net layers separated by alkaline earth atoms.

discovery stirred researchers. In that material, the CuO_2 planes do not have apical oxygen atoms, and are simply coordinated to the four neighbouring oxygens in the plane; the Cu-O bond lengths are longer than in the hole-doped materials. The superconductivity occurs in single CuO_2 planes separated by planes of Nd, Ce and O in the stacking sequence $\text{CuO}_2-(\text{Nd,Ce})-\text{O}-(\text{Nd,Ce})-\dots$

In 1988, Roth and collaborators discovered the compound which is now known as the 'all-layer phase' or the 'parent structure' of all copper-oxide superconductors³. It is based on the infinite stacking of CuO_2 planes separated only by the alkaline earth atoms, in the stacking sequence $\text{CuO}_2-\text{A}-\text{CuO}_2-\text{A}-\dots$ (see figure). Remarkably, at atmospheric pressure, that compound can be made only by mixing the small alkaline earth Ca with the medium alkaline earth Sr, at the one single composition, $\text{Ca}_{0.84}\text{Sr}_{0.16}\text{CuO}_2$, no doubt owing to the requirements of a delicate balance between the sizes of the CuO_2 plane and the alkaline earth layer. The discovery immediately triggered attempts to induce superconductivity in this simplest of all possible planar copper oxides, all of which had failed until now.

to be susceptible to electron doping. The successful execution of the electron doping required a detailed knowledge of previous results and a fancy synthetic route. Takano and coworkers, apparently working independently, have also just reported superconductivity in the all-layer compound, based on $(\text{Ba},\text{Sr})\text{CuO}_2$ prepared at high pressures⁵, although it is less clear just how the CuO_2 planes have been doped. Both materials may well turn out to be electron doped.

On the timescale of research in this field, an eternity separated the discovery of the first electron-doped copper oxide superconductor, $\text{Nd}_{2-y}\text{Ce}_y\text{CuO}_4$ and the second by Smith *et al.* — a testament to the difficulty of the materials since involved. Although $\text{Nd}_{2-y}\text{Ce}_y\text{CuO}_4$ does not appear to be the object of intensive research at present, there are still in my mind many questions left

1. Smith, M. G., Manthiram, A., Zhou, J., Goodenough, J. B. & Markert, J. T. *Nature* **351**, 549-551 (1991).
2. Tokura, Y., Takagi, H. & Uchida, S. *Nature* **337**, 345-347 (1989).
3. Siegrist, T., Zahurak, S. M., Murphy, D. W. & Roth, R. S. *Nature* **334**, 231-232 (1988).
4. Takano, M., Takeda, Y., Okada, H., Miyamoto, M. & Kusaka, T. *Physica C* **159**, 375-378 (1989).
5. Takano, M. *et al.* *Physica C* (in the press).

unanswered. What exactly, for example, is the crystal structure of the superconducting phase? Also, although that material and the new one appear clearly to be electron doped, there seem to be some positively charged current carriers in the materials which make physical property measurements ambiguous. One can, of course, hope that the study of the new electron-doped superconductor $\text{Sr}_{1-x}\text{Nd}_x\text{CuO}_2$ (and also $(\text{Ba},\text{Sr})\text{CuO}_2$) will

help to clear up some of the difficulties. Nature, however, never puts all her eggs in one basket: the possible simplicity of all-layer electron-doped copper-oxide superconductors will no doubt be partly offset by the fact that their relatively exotic synthesis will limit their availability for study. □

Robert J. Cava is at AT&T Bell Laboratories, Murray Hill, New Jersey 07974, USA.

GENETICS

Sweet mice, sugar daddies

Philip Avner

THE hunt for the genetic mechanism underlying diabetes will take on a fresh intensity with the observations by Todd and colleagues announced on page 542 of this issue¹. The authors have pinned down the chromosomal location of two separate genes, *Idd-3* and *Idd-4*, which confer susceptibility to type 1 (insulin-dependent) diabetes in mice; notably, both genes lie outside the major histocompatibility complex, whose role in autoimmune disease has already been established. But the significance of the work goes further, in that the technique employed by Todd *et al.* is one which could well be applied successfully to other complex diseases.

Type 1 diabetes (see box) belongs to a group of complex disease conditions that

includes atherosclerosis, hypertension, certain types of cancer and schizophrenia. We have only limited knowledge of the genetic basis of susceptibility to these diseases², but all of them are thought to be multifactorial — that is, to have both genetic and environmental components. All, moreover, are probably under polygenic control, with disease susceptibility resulting from the cumulative and interactive effect of several distinct genes.

The difficulty of studying such diseases in humans is in part due to their clinical and genetic heterogeneity, in part to the availability and informativeness of suitable cases. For some research groups the answer has been to develop animal models, the unspoken hope being that, at least in part, the same genes will control disease susceptibility

in the animal model and in humans. In the long run, then, such an approach could enable targeted genetic studies to be undertaken in humans, based on our knowledge of the homologies between the genetic maps of humans and the animal species in question. Given the proven role of the major histocompatibility complex (MHC) in both human and murine type 1 diabetes, there are grounds for thinking this approach is not purely wishful thinking. Animal models for polygenic multifactorial diseases include mouse models for certain types of spontaneous cancer³, a rat model for spontaneous hypertension, and the well-characterized BB rat and nonobese diabetic (NOD) mouse which provide productive analogues for probing the genetic and immunological basis of human type 1 diabetes (NOD mice spontaneously develop type 1 diabetes).

Studies on crosses involving mouse inbred strains, such as the NOD mice used in the study of Todd *et al.*, have in the past been severely hampered by the lack of informative genetic markers. Indeed, the main innovation of Todd and his colleagues has been to identify and map systematically a large number of simple repeat sequence microsatellite polymorphisms which provide the framework for subsequent exclusion mapping.

If the 1980s were the decade of *Mus spretus* — whose use in conjunction with restriction fragment length polymorphisms revolutionized mouse linkage analysis, and made the mouse a formidably efficient sys-

The Immunology of diabetes mellitus

DIABETES mellitus is a common pathological state defined by hyperglycaemia (high levels of blood glucose); it affects 2–4 per cent of the population in industrialized countries and may arise through various mechanisms. The two most frequent types of diabetes are insulin-dependent diabetes mellitus (type 1) and the non-insulin-dependent form (type 2). Both are chronic diseases. Type 1 usually appears abruptly as a hyperglycaemic syndrome in young people; type 2 is seen in older, frequently overweight, people. It has a more progressive onset and a strong genetic basis, and is associated both with dysfunction of the insulin-secreting cells and peripheral resistance to insulin.

Type 1 diabetes, by contrast, is the direct consequence of destruction of insulin-secreting (β) cells within the islets of Langerhans in the pancreas. Its incidence shows a north-south geographical gradient reaching 40 cases per 100,000 people in Scandinavia. Development of the disease is a multifactorial process in which it seems that environmental factors may trigger an autoimmune reaction against β cells in genetically predisposed people (the genetic element is estimated at 30–40 per cent). The only available treatment is replacement therapy involving daily injections of insulin.

The frequent occurrence of type 1 diabetes in association with other immune disorders, the presence of insulitis (infiltration

of the islets of Langerhans by lymphocytes) and the detection of islet cell antibodies at clinical onset of diabetes were all factors that first suggested that an immune process is involved. Type 1 diabetes is moreover associated with genetic markers that directly point to the immune system, genes in the HLA-D region on the short arm of chromosome 6 being associated with susceptibility to the disease. Studies of animal models, and therapeutic trials in humans, strongly implicate the involvement of T cells in particular. For instance, in humans islet β -cell destruction is transiently blocked by immunosuppressive drugs which target T cells. Other evidence for autoreactive T cells has come from the experimental transfer of diabetes, immunointervention using anti-T cell or anti-class II monoclonal antibodies in animals and, most importantly, the cloning of islet-cell-specific T cells.

A primary defect of autoantigen expression by β cells, defective autoantigen presentation, abnormal T-cell regulation of autoreactive T-cell clones and biased selection of the T-cell repertoire during ontogeny have all been invoked as allowing the development of anti-islet autoimmunity. Diabetes susceptibility genes, such as those mapped by Todd (see main article), could intervene in any one of these events. But it will be difficult to understand the breakdown of self-tolerance, and the triggering role of environmental factors, before

getting a clear picture of the mechanisms involved in tolerance to peripheral self-antigens. Studies of tolerance in transgenic animals should help in understanding how tolerance breakdown develops.

Knowledge of the interaction between autoreactive T cells and antigen-presenting cells, which is central to the activation of the anti-islet immune process, stems almost entirely from experiments in the BB rat and the nonobese diabetic mouse (see main article). This key interaction needs to be defined at the molecular level. The first hint in this direction was the characterization of class II MHC antigens associated with susceptibility haplotypes. If antigen-presenting molecules actually explain part of diabetes susceptibility due to the MHC, one may infer that a unique peptide initiates the activation of autoreactive T cells. Of course, T-cell clones will have to be extensively studied to define T-cell receptor usage by autoreactive cells and to characterize autoantigenic peptides and their role in initiating autoimmunity.

Characterization of non-MHC genes, such as those mapped by Todd *et al.*, and of their products and their effects either on the target cell or the immune system, are among the next steps to be taken.

Christian Boitard and Philip Avner

Christian Boitard is in INSERM U25, Hôpital Necker, Paris, France. Philip Avner is at the Institut Pasteur (see over).

Water at fault

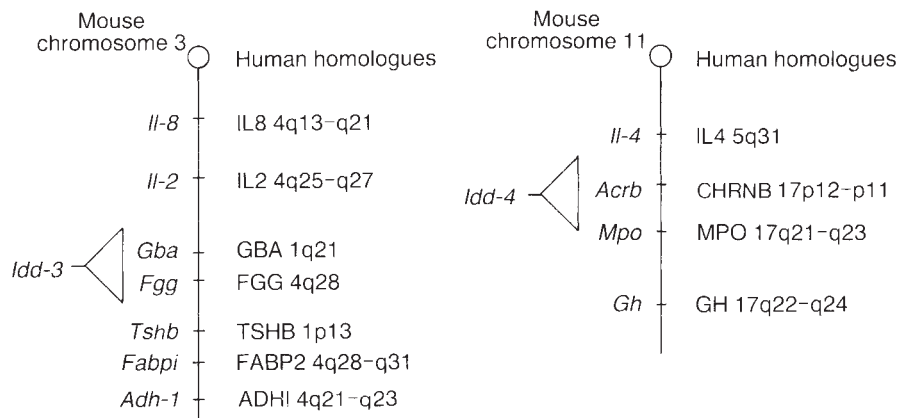
WATER is a crucial factor that can determine whether a crustal fault creeps gently or slips violently, M. L. Blandpied *et al.* find from laboratory experiments (*Geophys. Res. Lett.* **18**, 609–612; 1991). The distinction between smooth and unsteady slipping rests on whether the friction along the fault increases or decreases once movement starts. Laboratory tests in the past either used dry rocks or wet rocks, but not at realistically high temperatures. In the new experiments on water-bearing granite, Blandpied and colleagues find that stick-slipping motion occurs at just the temperatures (100–350°C) found at the depths at which seismic activity is recorded – between 1 and 15 km. At greater depths, the authors suggest, water allows the fault to deform smoothly, transmitting stresses from the convecting mantle which are ultimately responsible for the crustal seismicity.

Packing up

A COMMON view of the basis of the globular protein fold has been put to the test by M. J. Behe *et al.*, and found wanting (*Proc. natn. Acad. Sci. U.S.A.* **88**, 4195–4199; 1991). That view is that evolutionary R and D has resulted in amino-acid sequences that allow tight and highly specific packing of side-chains in a hydrophobic core, the core largely determining the fold. Behe and colleagues have searched for jigsaw-like, steric specificity in such tight-packed interiors and have found none. They conclude that the requirements for packing are not so restrictive after all, and that whereas the conformation of the amino-acid backbone indeed determines the packing, the packing in no sense dictates the fold. Does this mean that anyone will soon be able to design his own globular protein?

In a flash

CONCERNS about satellite debris in space are highlighted by notes in *Icarus* by P. D. Maley and by R. H. Rast (**90**, 326–327 & 328–329; 1991). 'Lunar transient phenomena' of many sorts have been reported over the years and one in particular – a brief, bright flash photographed near the edge of the Moon's crescent on 23 May 1985 – has been a source of great interest. Maley and Rast both conclude that the flash had nothing to do with the Moon. Rather, tumbling debris just happened to reflect sunlight at the wrong moment. Although the authors disagree over which piece of debris was responsible, the moral remains that astronomy will become harder as space becomes more polluted.



Known genes on mice chromosomes 3 and 11, to which Todd *et al.*¹ have localized the newly identified genes that confer susceptibility to type 1 diabetes. The suspected position of the two genes, *Idd-3* and *Idd-4*, is shown, as are the human regions homologous to the known genes. For example, the gene for interleukin-2 (*Il-2*) on mouse 3 has its homologue on the long arm (q) of human chromosome 4 on bands q25 to q27.

tem for genome mapping – the early 1990s look set to be the years of the microsatellite. Some 200 or so mouse microsatellites have now been identified⁴ and their use and that of alternative systems of analysis, such as mini-satellite fingerprinting⁵, will eventually allow crosses between classical inbred mouse strains to be almost as highly informative as those involving wild mouse strains such as *Mus spretus*. In this perspective, their use will inevitably lead to renewed interest in recombinant inbred strains established from laboratory mouse strains.

Todd and colleagues¹ report their application of an exclusion-mapping approach, based on these microsatellite sequences, to establish the chromosomal location of *Idd-3* and *Idd-4* to chromosomes 3 and 11, respectively, of the NOD mouse. These two loci are distinct from the MHC-linked loci that have already been identified in both humans and mice, which implies that susceptibility to murine type 1 diabetes is controlled by at least three unlinked loci. Although one previous gene association could not be confirmed⁶, as a minimal estimate this number is in agreement with previous studies using similar inbred strains. The implication of the present paper is that as the number of mice examined and the density of markers used in the exclusion mapping increase, further significant gene associations are likely to be identified using the same inbred-strain combination.

Important as the localization of *Idd-3* and *Idd-4* in the mouse is for diabetes research in both humans and mice, possibly of more fundamental consequence are Todd and colleagues' observations of the complexity of the genetic control of susceptibility to diabetes. For example, the *Idd-4* association with the disease is seen only in the youngest age-of-onset category of spontaneous diabetic animals. Moreover, neither the *Idd-4* nor the *Idd-3* locus shows behaviour compatible with a fully recessive gene model, and to this extent they rejoin and reinforce previous data obtained for the MHC-associated *Idd-1* locus. Here, the take-home message is

that although it is too soon to know what is really happening, complicated, allele-specific interactions at unlinked loci are probably at work.

Another finding is that *Idd-3*, but not *Idd-4*, is associated with insulinitis. This opens the possibility that different stages in the progression of type 1 diabetes will prove to be under the control of distinct genes or gene arrays. Progress here will obviously depend not only on improved genetic analysis but also on the application of more precise methods than classical histology for classifying the course of progression to diabetes.

A great deal of attention will now centre on the precise mapping of the gene associations reported by Todd *et al.*, and on a search for similar significant associations on the human chromosome homologues (17 for *Idd-4*, and 1 or 4 for *Idd-3*). But other questions involving the mechanism of gene control are likely to come to the fore in the mouse. Investigations of whether diabetes susceptibility in NOD mice is under the control of many more than three genes, with the identification of these other loci depending on the strain/cross combination examined⁶, are likely to be stimulated not only by the suggestion that the particular parent mouse used in the cross described by Todd *et al.* may contribute genes for resistance to diabetes, but also by the genetic analysis of crosses of NOD mice with other mouse strains. Before we know which, if any, of the non-MHC-linked loci so defined in humans and in mice are 'core' loci which must carry alleles for diabetes to develop, and learn how the multiple loci so defined interact, things are likely to become a lot more complicated. □

Philip Avner is at the Institut Pasteur, 28 rue du Dr Roux, 76724 Paris Cedex 16, France.

1. Todd, G. A. *et al.* *Nature* **351**, 542–547 (1991).
2. Weatherall, D. *et al.* *Molecular Approaches to Polygenic Diseases* Ciba Fdn Symp. Vol. 130 (Wiley, 1987).
3. Moen, C. J. A. *et al.* *Mammalian Genome* (in the press).
4. Hearne, C. M. *Mammalian Genome* (in the press).
5. Julier, C. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **87**, 4585–4589 (1990).
6. Prochazka, M., Leiter, E. H., Serreze, D. V. & Coleman, D. L. *Science* **237**, 286–289, (1987).

Unnatural selection

Christopher Foot and Andrew Steane

A NEW record-low temperature has been achieved in laser-manipulated sodium atoms in an experiment by S. Chu and coworkers at Stanford University (M. Kasevich *et al.* *Phys. Rev. Lett.* **66**, 2297–2300; 1991). Laser light selects atoms with a particular velocity from a sodium beam to define a new beam in which the velocity spread along one direction is only $270 \mu\text{m s}^{-1}$, corresponding to a one-dimensional 'temperature' of only 24 picokelvin. With such a precisely defined

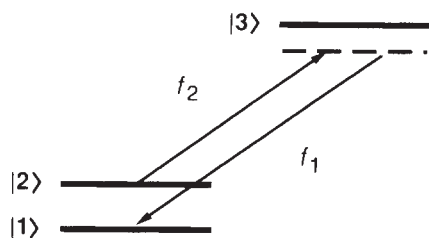


FIG. 1 The Raman transition used in the experiment transfers atoms between the states $|1\rangle$ and $|2\rangle$ — levels in the ground state of sodium. Laser light at frequencies f_1 and f_2 drives a coherent process of absorption followed by stimulated emission which goes via a virtual state close to the excited state of the atom, $|3\rangle$. The narrow linewidth of the transition gives Chu and colleagues great precision in the velocity-selection process. (From Kasevich *et al.*)

atomic beam, the authors proceeded to develop a matter-wave interferometer, exploiting the wave-particle duality of quantum mechanics.

The new experiment is the latest in a series of developments in which light forces have been used to manipulate and cool beams or clouds of atoms. The velocity-selection method achieves a reduction in velocity spread equivalent to a temperature decrease of six orders of magnitude when applied to atoms that have already been laser-cooled close to the limits of existing techniques. But this is not laser cooling in the strictest sense, as that involves the modification and reduction of atomic velocities by light forces to cool a whole vapour. In the new experiment, a spectral transition of narrow linewidth is used to select a subset of the atoms, about 1 in 1,000, from the original beam.

The spectral transition used is a Raman transition between two levels of the atomic ground state excited by laser radiation at two frequencies (Fig. 1). The transition is resonant when the difference between the frequencies of the light is approximately equal to the difference in frequency of the two atomic levels, Δf (the hyperfine-structure splitting in the ground state of sodium). Atomic motion tunes the exact frequency f through the Doppler effect: an atom moving at velocity v away from a light source sees an effective frequency of $f(1 - v/c)$, where c is

the speed of light. Thus when the laser beams at the two frequencies, f_1 and f_2 are counter-propagating, the Raman transition is resonant only for atoms within a narrow range of velocities about the velocity which fulfils the condition

$$f_1 - f_2 = \Delta f + \frac{v}{c} (f_1 + f_2)$$

These atoms are transferred from the lower to the upper of the levels and so separated from the other atoms. The advantages of the Raman transition are apparent from this equation. It is extremely sensitive to the atomic velocity, which is multiplied by the sum of the optical frequencies. Also, the selected velocity is determined by the frequency difference $f_1 - f_2$, and this is very stable as the two frequencies are produced by modulating a single original laser beam — the difference between the resultant 'sidebands' is determined by the applied modulation frequency from a stable microwave source.

At the extremely high precision of the experiment, even the small vibrations of the mirrors cause a significant change in the difference frequency, which has to be suppressed using electronic feedback. Ultimately, the linewidth of the Raman transition, and so the cooling limit, is determined by the Fourier-transform limit given by the reciprocal of the time for which the atoms interact with the laser radiation: the use of pre-cooled atoms ensures not only that there is a useful number of atoms to be selected from, but also that this interaction time is sufficiently long to allow a precise transition frequency.

The lowest temperature attainable by conventional laser cooling is limited by photon recoil. Each time a photon from the laser is absorbed by an atom, it imparts a momentum kick, the average effect of many such kicks being a force on the atom directed along the laser beam. This radiation force can cool atoms, but there is unavoidable heating produced by the recoil from the photons spontaneously emitted by the atom in all directions. When subjected to these random kicks, the atom's velocity cannot be reduced below a few times the recoil velocity from a single photon. But the coherent Raman transition does not involve spontaneous emission. Rather, one photon is absorbed and another radiated by stimulated emission in the opposite direction, to change the atomic momentum by twice the photon momentum, but without any spreading of the atomic velocity. (Note that a term equal to the recoil energy from two photons should be included on the right-hand side of the equation for the Raman resonance.)

Atoms colder than the recoil limit along one-dimension have been obtained previously by A. Aspect and colleagues (*Phys. Rev. Lett.* **61**, 826–829; 1988) at the Ecole

Normale Supérieure in Paris. Their experiment uses velocity-selective coherent population trapping, which is related in some ways to the new Stanford work. Both methods use Raman transitions excited in counter-propagating beams to select the transverse velocity of atoms, but the mechanisms differ markedly. The Parisian work used a so-called 'dark resonance' in the coherent transitions between two levels in metastable helium with the same total angular momentum ($J=1$). For these levels there is a superposition of the sublevels for which the quantum-mechanical absorption amplitudes destructively interfere. An atom that falls into this superposition state will remain there because it no longer interacts with the laser light. The trapping of atoms in this state only occurs for a narrow range of velocities as the interference of the amplitudes requires that the Raman transition be resonant — this criterion is similar to the equation above with $\Delta f=0$.

One important aspect of the coherent population trapping is that it is a true laser-cooling technique not just velocity selection. Atoms with finite transverse velocities, which are not initially trapped, absorb laser light and recoil from absorbed and spontaneously emitted photons until their velocity is just right for falling into the dark resonance. In this way an appreciable fraction of the atoms can be put into the trapping state. The method circumvents the recoil

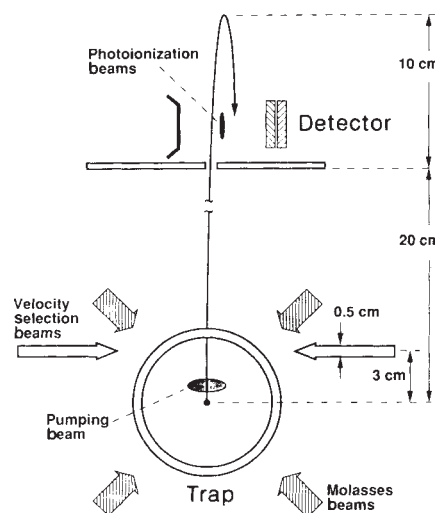


FIG. 2 Chu and colleagues' apparatus. Optical molasses (with six beams, two not shown) cool sodium atoms in a magneto-optical trap. Subsequently ejected, the atoms form a fountain from which some are velocity selected. See text for details. (From Kasevich *et al.*)

limit, because in the final state the atoms do not scatter photons, and the final transverse velocity of the metastable helium atoms corresponded to a temperature of 2 microkelvin.

The latest work at Stanford relies on several other laser techniques to cool the atoms before velocity selection. A beam of sodium atoms is first brought to a near halt by the

radiation force from a counterpropagating laser beam. The atoms are cooled further by a configuration of three orthogonal standing waves (Fig. 2) known as optical molasses, owing to the way the light exerts a damping force on the atoms so that they move like particles stuck in a viscous fluid. To build up a high density of atoms, a quadrupole magnetic field from a pair of coils with opposed currents was used in conjunction with the molasses laser beams to form a magneto-optical trap. The atoms are confined by an imbalance in radiation forces resulting from the Zeeman effect on the atomic energy levels (not by a direct magnetic force). Once a large number of cold atoms has accumulated at the centre of the trap, the current is turned off and the atoms are launched upwards using the light forces to form an atomic fountain; they travel for 30 cm before turning round and falling back down under gravity.

On their upward journey the atoms pass through a laser beam, which optically pumps them all into a single hyperfine level of the ground state before they enter the pair of counterpropagating beams that perform the velocity selection. The resulting velocity distribution is measured by passing the atoms through a small slit in a liquid-nitrogen-cooled shield into a region in which the atoms in the upper hyperfine level can be detected by counting the ions from laser-photoionization. The transverse width of the atom beam was that of the slit — there was no observable spreading. The upper limit for the transverse velocity was thus 1 mm s^{-1} . The value of 0.27 mm s^{-1} quoted by Chu and colleagues was inferred from the measured line-width of the Raman transition — it is difficult to measure velocities this small. In fact the authors showed that selection and detection of atoms across a range of initial velocities determines the initial distribution more accurately than the usual time-of-flight methods.

Kasevich and Chu have devised, in work yet to be published, an atom interferometer, analogous to an optical interferometer, that relies on the wave-like properties of atoms. The de Broglie wavelength of the atoms is $54 \text{ }\mu\text{m}$, projected along the laser beams. Consider atoms which start in the lower level and with momentum p along the laser beams; we denote these by $|1, p\rangle$. A first Raman pulse creates a superposition of the initial state $|1, p\rangle$ and $|2, p+2\hbar/\lambda\rangle$, in which the atom is in level 2 with additional momentum equal to twice the photon momentum. These wavepackets will move apart and become separated by as much as 2.4 mm in this initial experiment. A second Raman pulse, twice as long as the first, swaps over the states of the two packets so that they move back together. A third Raman pulse, with the same duration as the first, recombines the amplitudes — the final atomic state is a mixture of the two levels that depends on the relative phase difference of the two paths (and the laser fields). Kasevich and Chu based their matter-wave

interferometer on this principle, but with the Raman beams vertical, to determine the gravitational acceleration of sodium atoms to a few parts per million.

Matter-wave interferometers have been made before using electrons and neutrons. But experiments with atoms are only now coming to fruition. Two groups (O. Carnal and J. Mlynek *Phys. Rev. Lett.* **66**, 2689–2692; and D. W. Keith *et al.* 2693–2696; 1991) have just reported suc-

PARTICLE ACCELERATORS

Focusing for free

Roger Evans

WHEN most people think about focusing, whether for light waves or for charged particles, the mental picture is of carefully designed and exquisitely manufactured components of the closest possible tolerances. It comes as something of a surprise to find, as H. Nakanishi *et al.* have (*Phys. Rev. Lett.* **66**, 1870–1873; 1991), that nature provides a near ideal focusing system for charged particles in nothing more than a puff of ionized gas.

The possible advantages of variants of this 'plasma lens' concept have been widely discussed by those operating particle accelerators in their quest for ever-higher-energy machines. Future multi-teraelectronvolt accelerators will have to focus their colliding beams down to almost atomic dimensions if interesting collisions are to occur at a sufficient rate to satisfy the experimenters.

Conventional quadrupole magnets focus beams in only one of the two transverse planes. The ability to focus charged particles in both planes simultaneously requires there to be a charge density or current in the focusing region. Controlling the uniformity of the necessary plasma has usually been thought to be insurmountable given the propensity of plasmas to collapse via a wide range of instabilities. But the new experiments by Nakanishi *et al.* show that the charge distribution required to focus the beam can be produced by the beam itself and is in this sense 'free', only a background plasma has to be provided. The mechanism is straightforward and depends on the fact that the particles in a highly relativistic beam in a vacuum experience competing focusing and defocusing forces: the electrostatic repulsion between the charged constituents is defocusing, but the magnetic attraction arising from their 'parallel currents' cancels out the repulsion to a factor depending on how close to the speed of light the beam travels.

When the beam passes through a background plasma, the plasma electrons will move in or out in an attempt to cancel the space charge, leaving the attractive electromagnetic force to pinch the beam particles. If the beam density is uniform the attractive force increases linearly with beam radius so that it is a near ideal focusing lens. Experi-

ment in making atom interferometers using fine slits and diffraction gratings, respectively, to observe the interference of atomic wavepackets. Interferometers like these and that at Stanford based on laser-cooled atoms could be used in new tests of the remarkable subtleties of quantum mechanics. □

Christopher Foot and Andrew Steane are in the Clarendon Laboratory, University of Oxford, Parks Road, Oxford OX1 3PU, UK.

ments with 18-megaelectronvolt electrons at the University of Tokyo have confirmed this effect and shown that the quality of focusing is even better than a perfect lens. This paradoxical result is tentatively ascribed to a coupling of the transverse and longitudinal motions so that the all important 'phase-space volume' (a dynamic notion) is conserved.

Unfortunately, just as in economics, there are no 'free lunches'. This free focusing has some drawbacks. In its passive form, as described above, the strength of the focusing action increases as the beam passes through the plasma, the tail is focused much more strongly than the head which is not very practical for a colliding-beam system. The focusing is symmetrical for positive and negative particles only if the beam density is much less than the plasma density, in which case one might worry about scattering from the plasma particles. A practical disadvantage is that the focusing lens appears to move as the input beam moves which makes the alignment tolerances much more critical for two beams to collide at the same point.

Novel ideas such as this plasma lens or plasma-based accelerators have been part of the imaginative schemes put forward for particle accelerators to advance to the next generation. Although the new ideas always have some highly attractive features it is surprising just how resilient and ingenious are the designers of conventional accelerators and so far all serious designs have been based on radiofrequency acceleration and magnets for bending and focusing. The plasma lens is unlikely to be used as the sole focusing element but its cheapness and simplicity make it the ideal way to enhance existing systems. A passive plasma lens close to the focus of two colliding beams may well provide a useful increase in the luminosity at the intersection for very little cost. But its use in circular machines is sadly limited: as the action is not reversible, the focused beam cannot be recollimated; and the accumulation of scattering from plasma particles will eventually destroy the beam quality. □

Roger Evans is at the Rutherford Appleton Laboratory, Didcot OX11 0QX, UK.

Oh to be twenty seven again

Peter Parham

DISCRIMINATION by T cells of friend and foe is based on displays of peptides derived from intracellular and extracellular proteins. By its generality this mechanism gets antigen-presenting molecules of the immune system involved with almost everything else in the cell. In turn, the pursuit of antigen processing and presentation forces immunologists to sail from the coastal comforts of the plasma membrane and explore the cell's interior.

This coming about was reflected at a workshop last month*, where interest centred on the natural peptides bound by antigen-presenting molecules, the molecular mechanisms of peptide binding and the cellular compartments in which it takes place. In this setting T cells were a generic readout of function, the thymus was rarely mentioned and cellular interactions were, well, almost infra dig (D. Y. Loh, Washington University). Knowledge of the primary and three-dimensional structures of bound peptides *in vivo* is rapidly advancing, and a newly discovered protein, involved in the assembly of class I antigen-presenting molecules, is firmly on the scene.

Early running in this area was set by studies on the T-cell response to exogenously supplied antigens. Such antigens are mainly presented by class II major histocompatibility complex (MHC) molecules and it was for class II that the dependence of T-cell recognition on binding of degraded antigen fragments to MHC molecules was established. This principle was subsequently shown to extend to class I molecules, but with nuances in mechanism that favour the presentation of intracellular antigens. One difference is that class I molecules bind peptides near their site of synthesis in the endoplasmic reticulum (ER), whereas class II molecules are targeted to compartments of the endocytic pathway by the invariant chain. This trimeric molecule binds class II molecules in the ER, takes them to endosomal and lysosomal compartments for peptide binding, and is itself destroyed in the process (P. Cresswell, Duke University).

A search has been made for possible chaperonin or targeting molecules that interact with class I molecules during their biogenesis. From crosslinking experiments a good case can now be made for transient association of class I heavy chains with a polypeptide of relative molecular mass 88,000 (p88) (D. B. Williams, University of Toronto). In the ER, each newly synthesized class I heavy chain rapidly forms a 1:1 complex with p88. This association does not require prior association with β_2 -microglobulin (the second and invariant polypeptide of the class I molecule) and is not perturbed by its subsequent

binding. The signal that triggers release of class I molecules from the p88 complex is not yet known, but dissociation correlates precisely with the acquisition of terminal sugars in the medial Golgi. These kinetics, combined with experiments implicating the class I transmembrane region in p88 interactions, imply that p88 does not directly facilitate folding of class I molecules but acts to retain them in the ER, perhaps until they are loaded with peptide. Another possibility, based on similarities in molecular weight and protease sensitivity, is that p88 is the product of one of the putative peptide transporter genes encoded by the MHC (see my News and Views article in *Nature* 351, 271–272; 1991). Intimate association between peptide pump and nascent class I molecules could pinpoint delivery of peptides and in this manner facilitate the correct assembly of class I molecules.

Determination of the three-dimensional structure of a third human class I antigen-presenting molecule, HLA-B27, is nearing completion (D. C. Wiley, Harvard University). Although the product of a different genetic locus, HLA-B27 has a peptide backbone that is essentially the same as that shared by HLA-A2 and HLA-Aw68. This augurs well for a structural generality, because by physicochemical criteria A2 and B27 are at opposite ends of the spectrum of HLA-A,B diversity. The special contribution of this third class I structure is improved resolution in the 'extra density' of the antigen-binding groove and proof it is a peptide. Whether it be due to the peculiarities of B27, better crystals or third-time luck, this eagerly awaited image of an endogenous peptide is no disappointment. The peptide appears as an extended chain of nine amino acids, not the α helix of popular prediction, with its ends tethered at opposite ends of the groove. Side chains of some amino acids extend deep into specificity pockets formed by variable positions of the class I heavy chain, others point up with the possibility for direct contact with the T-cell receptor. Such contacts, however, remain a three-dimensional mystery, as illustrated by systematic analysis of the T-cell receptors interacting with mutant class I molecules and a malarial peptide (P. Kourilsky, Institut Pasteur).

Until last year, almost all studies of the epitopes presented by MHC molecules to T cells used synthetic peptides. From the results for both class I and class II molecules it seemed that optimal epitopes were about 12–15 amino acids in length. Now, characterization of the 'natural' peptides bound by class I molecules within cells shows they are both smaller and more precisely defined than expected (G. van Bleek, Albert Einstein College of Medicine; H.-G. Rammensee,

Max-Planck-Institut für Biologie, Tübingen). Nine amino acids is the current favourite and this number, which is consistent with the crystallography, echoed throughout the workshop. From sequencing both mixtures and individual peptides bound by particular MHC molecules, conserved motifs of 'anchoring residues' — one at the C terminus and another close to the N terminus — that distinguish sets of binding peptides for different class I molecules are being identified. None of these peptides seems to be longer than nine amino acids. This convergence of crystallography and biochemistry sets a new paradigm for class I epitopes. A key question is whether the same will be true for class II molecules, as there are many suggestions that their binding sites are more accessible to experimentally offered peptide, more tolerant of peptide length and able to undergo reversible peptide exchange as they recycle through the cell (L. Adorini, Sandoz Pharma, Basel; M. M. Davis, Stanford University; B. Pernis, Columbia University).

The unexpectedly short length of naturally presented peptides inevitably raises questions about the biological activities of longer synthetic peptides. Indeed, these 'defined reagents' are not always as advertised. For example, an epitope of Sendai virus was defined using a panel of synthetic 'overlapping' 12-residue peptides derived from the nucleoprotein sequence. Peptides from a single region of the protein were presented to T cells, thereby mapping the epitope. But the active principle in peptide preparations that actually bound to the class I molecules was a very minor constituent. This contaminant was nine amino acids in length, differing from the expected sequence by the absence of the three N-terminal amino acids — the last ones to be incorporated in the synthesis (C. J. Melief and H. L. Ploegh, Netherlands Cancer Institute).

In reevaluation of assays of peptide binding and presentation, smaller peptides are often better. For peptide-driven association of class I molecules in cell lysates (A. Townsend, University of Oxford), for inhibition of peptide binding to the class I molecules of viable cells (B. Barber, University of Toronto) and for reconstitution of HLA-A2 (D. Wiley) nine-residue peptides have more potent activities than their longer, previously advocated, homologues. Clearly, synthetic peptide preparations will now have to be scrutinized closely and, though the immediate prospect is tedious, there are benefits to this turn of events. Synthetic peptides are expensive, and, like yachts, the longer they are the more costly they become. When appropriate endogenous epitopes have been identified, the necessary peptides will not only be cheaper to make, but greatly increased in biological activity and ultimately longer lasting. □

Peter Parham is in the Departments of Cell Biology, and Microbiology and Immunology, Stanford University School of Medicine, Stanford, California 94305, USA.

*Molecular and Cellular Biology of Antigen Processing and Presentation, Cancer Research Institute, New York, 6–7 May, 1991.

Crying wolf in North America

John L. Gittleman and Stuart L. Pimm

IN A world in which natural habitats are shrinking alarmingly, which species shall be saved? The task of classifying species falls of course to taxonomists, and among the priorities for our limited Ark are often those plants and animals that enjoy taxonomic distinctiveness at high levels (species in monotypic genera or families, for example). Taxonomists also decide which populations are to be considered legitimate species — bad taxonomy can kill when distinct species are not afforded specific status¹. Wayne and Jenks, in their analysis of the specific status of the red wolf on page 565 of this issue², describe a high resolution way of distinguishing between closely related mammals. Their finding — that the red wolf is not distinct from the grey wolf or coyote — is however less important than the issues it highlights.

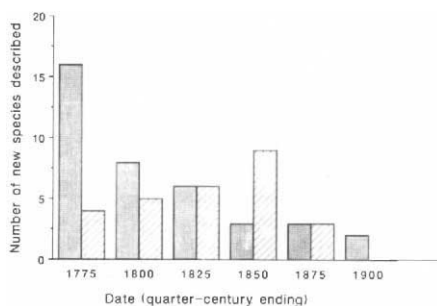
The red wolf has been the beneficiary of a substantial captive breeding and re-introduction campaign following its extinction in the wild. But its recognition as a species, *Canis rufus*, distinct from the closely related grey wolf (*Canis lupus*) and coyote (*Canis latrans*), has always been a matter of controversy. As a first indication of this controversy consider when the red wolf was first described — in 1851, by Audubon and Bachman in *The Quadrupeds of North America*, an unusually modern discovery when compared to both canids worldwide and all North American carnivores (see figure).

Why was a species of such large size, conspicuous predatory habits and widespread distribution first described at such a late date? Interestingly, the primary proponents³ of species status for the red wolf drew their conclusions less than two decades ago. They prefaced their analysis by stating: "... there is such a great degree of individual, geographic, sexual, and age variation within each species of North American *Canis*, and such wide overlap in most characters, that most of the following diagnosis is necessarily general and qualitative in nature" (p.1). Using multivariate techniques, only a few morphological characters out of over a dozen showed separation of *C. rufus* from *C. latrans* or *C. lupus*, and these characters seemed to be size-related and intermediate between those seen in the coyote and grey wolf⁴. This is not to say that morphologically almost identical species do not exist. Early studies with protein electrophoresis did not resolve any differences from the other *Canis*⁵ either. Nevertheless, carnivore taxonomists have continued to accord species status to *C. rufus*^{6,7}.

Wayne and Jenks's use of more modern techniques of assaying mitochondrial DNA (mtDNA) provides a more compelling appraisal of the red wolf's taxonomic status. First, they compared the mtDNA from a small captive breeding colony of red wolves with samples from 327 coyotes and 276 grey

wolves. Phylogenetic analysis shows the red wolf mtDNA genotype to be very like that of coyotes, and that it has particular similarity to the mtDNA of coyotes living in those areas from which the red wolf had disappeared.

The colony may not adequately represent the wild populations, however, because, as in other tiny populations, genetic variability



Dates of first description of Carnivora in North America (left) and Canidae worldwide (right) (from refs 5 and 6). The red wolf was described only comparatively recently, in 1851 by Audubon and Bachman (and was resurrected as a species in 1972, by Paradiso and Nowak³). Subsequent descriptions of species have been mostly of rare, insular forms, and in most cases their taxonomic status remains controversial.

can be lost quickly. Fortunately, between 1974 and 1976 blood samples were collected and stored from wild animals classified on morphological grounds as red wolves. Analyses of the mtDNA from the samples corroborated the previous results. Taken alone, these results raise serious doubts about the red wolf as a species.

But these modern animals may be the last hybridized gasp of a once viable species, which became so rare that the individuals had no choice but to mate with coyotes. So how should we resolve what the red wolf was really like? Definitive conclusions are made possible by the existence of a direct source of historical information, the DNA sequence of the cytochrome *b* gene from museum pelts taken from putative original red wolf populations in the southeastern United States. Again, the DNA sequence data show that grey wolves and coyotes are clearly distinct from each other, with the red wolf always being similar to either one or the other of these species. The elegance of Wayne and Jenks's study is in the combination of baseline morphological data with a two-pronged genetic analysis: evaluation of current genetic variation along with assessment of historical variation among species.

Various taxonomic issues are raised by this paper. First, species identification using morphological and molecular data in concert really does provide diagnosis and verification in systematics; ambiguity in the

morphological systematics of the red wolf indeed foreshadowed a similar result in mtDNA. Molecular systematics is only beginning to develop protocols for considering how much and what kind of differences in mtDNA are necessary to erect species status. (In the case of the red wolf, as in many other hybridized mammals, differentiation is non-existent so the problem is moot.) Second, when differences are detected, especially slight differences among closely related taxa, molecular studies must provide quantitative information about where the differences arise in the various classes of mtDNA (base substitution and length variation). Such information permits statistical analyses of the evolutionary differences between species that could not be easily resolved in the past.

For decision-making in conservation, there are ecological issues as well as taxonomic ones. Obviously, the areas richest in species will have to be identified. We are unlikely to arrive at tallies of species by counting one, two, three... in each area, for taxonomic catalogues will not be complete in the foreseeable future. Rather, ecological knowledge about patterns of species richness is essential. We must also expect to manage species within the fragments of once extensive habitats and, in particular, identify which species play 'keystone' roles and whose loss will cause further secondary extinctions. When selecting species for management, we must ask which are in immediate danger and, of course, understand why some species are more prone to extinction than others. And when all this ecological knowledge has been gathered, difficult policy choices will have to be addressed — which species we are likely to be able to save; which areas are easiest to protect with available resources; and the controversial topic of whether to emphasize saving species, or saving communities, or saving ecosystems⁸.

The loss of large predators has been particularly severe worldwide: their body size and position in the food chain makes them rare and so vulnerable to extinction (their perceived rivalry with humans even more so), and they play important functional roles at the top of food chains. Special steps should be taken to preserve them. The large geographical areas they require, combined with their public appeal, means they can demand large protected areas, where the many other species present could not. That appeal also permits us to learn about how to re-introduce species, lessons we can then apply to saving

1. May, R. M. *Nature* **347**, 149 (1990).
2. Wayne, R. K. & Jenks, S. M. *Nature* **351**, 565–568 (1991).
3. Paradiso, J. L. & Nowak, R. M. *Mammalian Species Account* No. 22 (1972).
4. Paradiso, J. L. *Mong. Univ. Kansas, Mus. Nat. Hist.* **6**, (1979).
5. Ferrell, R. E., Morizot, D. C., Horn, J. & Carley, C. J. *Biochem. Genet.* **18**, 39–49 (1980).
6. Honacki, J. H., Kinman, K. E. & Koepl, J. W. *Mammal Species of the World* (Assoc. Systematics Collections, Lawrence, Kansas, 1982).
7. Wozencraft, W. C. in *Carnivore Behavior, Ecology, and Evolution* (ed. Gittleman, J. L.) 569–593 (Cornell Univ. Press, Ithaca, New York, 1989).
8. Soulé, M. E. & L. S. Mills. *Oikos* (in the press).

less 'charismatic' species.

How does the re-introduction of the red wolf relate to these ecological reasons for protecting species? Information gathered by the captive breeding programmes about, for example, artificial insemination, may be valuable in saving other less controversial taxa. Likewise, the introductions of the red wolf to islands and other areas where the coyote is absent may provide information about how to re-establish similar species. From just which taxa we learn these lessons may be less important than the lessons themselves.

Wayne and Jenks's results on the red wolf suggest that here no species is being saved (and certainly nothing of great taxonomic

distinctness); that in areas where the coyote is already present, no new ecological role in the community is being created by the introduction of red wolves; and that any lessons about re-introducing species are likely to be obscured by hybridization with the coyote. No new reserves are being created for the red wolf. In this case, then, the re-introduction comes down to asking whether the red wolf's undeniable cuddliness is enough to warrant according it special attention. □

John L. Gittleman and Stuart L. Pimm are in the Department of Zoology, University of Tennessee, Knoxville, Tennessee 37996, USA.

The rift narrows

Rob Ellam

OVER the past few years, physical models of the melting processes that occur in the Earth's crust and mantle, producing the lavas erupted by modern volcanoes and the igneous rocks of the geological record, have reached new levels of sophistication. Igneous petrologists, who previously relied largely on chemical information to elucidate magmatic processes, have not been slow to embrace the physical approach, but some intriguing questions have emerged. One particular puzzle is the occurrence of magmatism where there is little tectonic extension and where there is no obvious relationship to anomalously hot plumes of mantle like those beneath oceanic hotspots such as Hawaii and Iceland. Previous models had difficulty in explaining even the modest amounts of basalt found in these areas, but on page 559 of this issue¹, Latin and Waters discuss refinements in the melting models that appear to explain these somewhat enigmatic magmatic events.

When tectonic processes stretch the rigid continental and oceanic plates (lithosphere), the underlying convecting mantle (asthenosphere) is forced to rise. If the pressure release is sufficient, the mantle will melt. To predict the amount of melting for a given degree of lithospheric extension requires knowledge of the mantle solidus curve, the line on a plot of pressure versus temperature that defines the initial occurrence of partial melt. In 1988, McKenzie and Bickle² used the results of experimental melting studies to define the solidus and thereby calculate the melt distribution beneath a mid-oceanic ridge. They concluded that for mantle of normal temperature, that is where there is no influence from a hot ascending mantle plume, melting is restricted to the upper 50 km or thereabouts of the asthenosphere.

In a continental rift, like the North Sea graben studied by Latin and Waters, the pre-rift lithospheric thickness probably approaches 120 km, so that the lithosphere would need to thin by a factor of about 2.5

(the β -factor) while stretching before the rising asthenosphere penetrates the maximum depth at which melting occurs (about 50 km). Not least because the amount of extension is critical in the assessment of the petroleum potential of a rifted basin, β -values tend to be well constrained in the North Sea. For the most stretched part β is only about 2, yet basalts from the Forties volcanic province indicate that quite substantial magmatism has occurred, apparently without the influence of elevated mantle temperature.

McKenzie and Bickle² themselves identified the uncertainties likely in laboratory experiments as an important restriction on their conclusions. Experimental charges are, by necessity, both very small and enclosed. Thus the experiments simulate batch melting, in which all the melt produced remains in equilibrium with the solid residue until it separates as a single batch. In the mantle, even very small melt fractions (less than 2 per cent by volume) are probably able to escape from a deforming matrix³, and a 'dynamic' melting model is perhaps more appropriate (see, for example, the recent paper by Elliot *et al.*⁴). The melt distribution generated by dynamic melting might differ significantly from the experimental prediction.

Now McKenzie and O'Nions⁵ have used an inversion technique to calculate melt distributions from rare-earth-element (REE) abundances in mid-ocean ridge basalts, assuming a Rayleigh melting process in which melt and matrix separate as soon as the melt is formed, and which therefore better approximates a dynamic melting process. Critically, melting commences at about 80 km depth, which corresponds to $\beta = 1.5$ for a normal initial lithospheric thickness. Using the new calculated melt distributions it is possible to have melt formation between 80 km and the base of the stretched lithosphere in low- β rifts such as the North Sea, without recourse to elevated mantle temperatures.

However, when Latin and Waters apply

the results of the REE inversion to the North Sea case, there is significant discrepancy between the relatively low light-REE concentrations predicted by the Rayleigh melting model and the light-REE-rich basalts of the North Sea. An additional source of light REE is required. Nephelinite and ultrapotassic magmas found towards the flanks of the graben apparently provide the additional REEs, and have suitable Nd isotope compositions to explain the low (close to bulk Earth) $^{143}\text{Nd}/^{144}\text{Nd}$ ratios of the Forties basalts compared with typical depleted mantle. The nephelinites are taken to originate in lithospheric mantle, not the asthenosphere, that has been metasomatized (altered by fluids). This melts in response to rifting because the presence of volatile-rich phases substantially reduces its solidus temperature.

The role of lithospheric material in magma generation is of great interest to geochemists because the mechanical boundary layer (approximately the upper 100 km) of the lithosphere is isolated from the homogenizing effects of asthenospheric convection. In the mechanical boundary layer, isotopic heterogeneities can evolve for long periods of time ($1-2 \times 10^9$ years) in response to ancient changes in trace-element ratios. For example, domains in the boundary layer with low Sm/Nd ratios will gradually evolve lower $^{143}\text{Nd}/^{144}\text{Nd}$ ratios than the average (higher Sm/Nd) asthenosphere. On melting, the mantle portion of the boundary layer yields melts of broadly basaltic composition, but with unusual isotopic signatures. A recent breakthrough has been to define rather precisely the lithospheric contribution to extensional magmatism. Thus in the North Sea, Latin and Waters demonstrate the involvement of magmas similar to locally occurring nephelinites, whereas Thompson *et al.*⁶ in northwest Colorado and our own work⁷ in the Karoo flood basalt province of southern Africa have identified ultra-potassic, lamproitic end-members also attributed to fusion of volatile-rich lithospheric mantle.

Even these few studies serve to illustrate the assortment of magmas that might be produced by partial fusion of enriched lithospheric mantle. Clearly the geochemical effects of interaction between asthenospheric and lithospheric melts will be very variable. An obvious goal for the future must be to relate those variations to the diverse metasomatic processes that apparently control the formation of the low-melting-temperature fraction of the mantle lithosphere. □

Rob Ellam is in the Isotope Geology Unit of the Scottish Universities Research and Reactor Centre, East Kilbride, Glasgow G75 0QU.

1. Latin, D. & Waters, F. G. *Nature* **351**, 559-562 (1991).
2. McKenzie, D. & Bickle, M. J. *J. Petrol.* **29**, 625-679 (1988).
3. McKenzie, D. P. *Earth planet. Sci. Lett.* **74**, 81-91 (1985).
4. Elliot, T. R. *et al.* *Nature* **351**, 201-206 (1991).
5. McKenzie, D. P. & O'Nions, R. K. *J. Petrol.* (in the press).
6. Thompson, R. N. *et al.* *Earth planet. Sci. Lett.* **98**, 139-153 (1989).
7. Ellam, R. M. & Cox, K. G. *Earth planet. Sci. Lett.* (in the press).

Cuddling up in the dark

Michael P. Stryker

IT is known that neural activity is involved in the courtship by which afferent nerve terminals establish relations with appropriate target cells in some parts of the nervous system. But whether the relevant activity is that of the presynaptic afferent, the postsynaptic cell or some harmony between the two has remained in question — most methods of interfering with activity perturb it on both sides. Now, on page 568 of this issue¹, Hahm *et al.* provide compelling anatomical evidence that postsynaptic responses play a crucial role. This paper follows close behind another, in *Science*², that reveals great order in the patterns of retinal activity occurring in development before vision is possible.

In most mammals with good binocular vision, the connections from the eyes to the main visual relay nucleus, the lateral geniculate nucleus (LGN), are arranged in layers that each receive input from one eye or the other. In a number of species, geniculate layers or sublaminae are also specific for the centre type ('On' or 'Off', responding to lightness or darkness, respectively) of the innervating retinal ganglion cells³.

At the level of the single cell, the specificity of connections is even greater: about half of the geniculate cells receive their effective excitatory input from only a single retinal ganglion cell; and among cells that receive several inputs, the inputs always originate from the same eye and from retinal ganglion cells of the same centre type⁴. Anatomical studies have revealed that single geniculate cells actually do manage this feat of ensuring that their hundreds of synaptic contacts are made with only one or a few of the hundreds of retinal afferents with which their dendritic fields overlap⁵.

How are these synaptic partners chosen with such precision? A most attractive hypothesis has been that discharge activity of retinal ganglion cells is locally correlated among cells of the same centre type within each small region of the retina, so that the timing of neural activity provides a high-resolution signal of the eye of origin, neighbourhood relationships and centre-type of each retinal terminal at every point within a target structure. If early synapses were hypothesized to have the hebbian property that they increased in strength according to the correlation between pre- and postsynaptic activity, then such input activity could lead initially diffuse projections to refine into precise maps, layers and single inputs⁶.

In normal development, retinal inputs from the two eyes initially innervate the LGN without regard to layer. Later (during the last few weeks of fetal life in the cat) they lose connections to the inappropriate regions and vastly expand their synaptic terminal arbor (branching structure) in appropriate regions so as to form eye-specific layers⁷.

Although there is no visual stimulation during this period of life, retinal ganglion cells are active⁸ — they seem to discharge rhythmically, in waves of activity that propagate at various orientations across the retina². Blocking this pre-visual activity of the retinal nerve fibres (and probably also that of the geniculate cells) by infusions of tetrodotoxin into fetal brains prevents the eye-specific layers from forming⁹.

Hahm *et al.*¹ have now found that, in normal development, retinal afferent terminal arbors are progressively confined to single geniculate sublaminae. Locally or systemically applying blockers of the *N*-methyl-*D*-aspartate (NMDA) subtype of glutamate receptor, which mediates much of the input to the geniculate cells, would be expected to have little or no effect on retinal afferent activity, but it prevents the normal rearrangement of the retinal arbors¹. These new results are consistent with the operation of hebbian synapses in the normal development of retinogeniculate connections.

The NMDA receptor has attracted enormous interest, because its biophysical properties allow it to inject calcium into the postsynaptic cells specifically at synapses whose synaptic activation takes place in association with other activity that depolarizes the cell¹⁰. In the CA1 region of the hippocampus, calcium entry through NMDA receptors seems to be necessary, and perhaps sufficient, for the long-term potentiation model of adult synaptic plasticity¹¹. The NMDA receptors may play a similarly crucial part in amphibian visual-system development¹². In the mammalian visual system, both pre- and postsynaptic cells may be studied anatomically and physiologically *in vivo* and *in vitro*. Future studies of this system may reveal whether calcium entry through NMDA receptors, or some other consequence of harmonious pre- and postsynaptic activity, mediates the synaptic plasticity responsible for the development of precise synaptic connections. □

Michael P. Stryker is in the Department of Physiology and the Neuroscience Graduate Program, University of California, San Francisco, California 94143-0444, USA.

1. Hahm, J.-O., Langdon, R. B. & Sur, M. *Nature* **351**, 568–570 (1991).
2. Meister, M. *et al. Science* **252**, 939–943 (1991).
3. Stryker, M. P. & Zahs, K. R. *J. Neurosci.* **3**, 1943–1951 (1983).
4. Mastronarde, D. N. *J. Neurophysiol.* **57**, 357–413 (1987).
5. Hamos, J. E. *et al. J. comp. Neuro.* **259**, 165–192 (1987).
6. Stent, G. S. *Proc. natn. Acad. Sci. U.S.A.* **70**, 997–1001 (1973).
7. Shatz, C. J. *J. Neurobiol.* **21**, 197–211 (1990).
8. Galli, L. & Mafei, L. *Science* **242**, 90–91 (1988).
9. Sretavan, D. W., Shatz, C. J. & Stryker, M. P. *Nature* **336**, 468–471 (1988).
10. Nowak, L. *et al. Nature* **307**, 462–465 (1984).
11. Malenka, R. C. *et al. Science* **242**, 81–84 (1988).
12. Constantine-Paton, M. *et al. A. Rev. Neurosci.* **13**, 129–154 (1990).

Ghostly graphite

CARBON fibre is made by carbonizing a carefully tensioned polymeric fibre. The tension aligns the growing graphite molecules in their 'strong' direction, giving the resulting fibre great strength and stiffness. But graphite is essentially two-dimensional. An oriented carbon-sheet product should be stronger and stiffer still. So Daedalus is inventing one.

His first idea was simply to carbonize acrylic sheeting while stretching it in two dimensions. But he now plans to start from a foamed polymer such as polyurethane. DREADCO's chemical engineers are polymerizing and foaming very dilute solutions of such materials. The resulting products resemble ultra-fine soap-foam. As the solvent evaporates, the cell walls of the foam thin to the desired thickness.

To graphitize this material evenly, Daedalus heats it in a big microwave oven. Pyrolysis-gas and residual solvent vapour build up in the closed foam-cells, and each tiny film-facet as it chars is tensioned and oriented by the resulting expansion. As in all liquid-foam structures, these individual facets all meet at 120°. At this angle, graphite sheets should combine very firmly, in a sort of polymeric extension of the triptycene molecule. When the process has been properly optimized, the resulting 'foamed graphite' should be immensely strong and rigid for its weight.

DREADCO's foamed graphite will be available in many grades. Thick foams will give dense materials resembling matt-black wood, ideal for load-bearing components such as beams and planks. More attenuated foams, formed as thin sheets and faced with polymer film, will compete with aluminium sheet in stressed-skin structures such as aircraft and pressure-piping. The thinnest foam of all will give an open network of monolayer graphite. This tenuous and ghostly product, a sort of polymeric higher fullerene, will have a low bulk density and will be very difficult to see. Like that other expanded solid, silica aerogel, it will be mainly empty space. Even so, it will have surprising strength and rigidity, amazing thermal insulation properties and will conduct electricity in strange ways. Like activated charcoal, it should also absorb gas readily.

If that gas is oxygen, of course, the result will be combustible or even explosive. The thicker, structural varieties of foamed graphite should take up gas too slowly to be hazardous from this effect. Even so, Daedalus is planning an autodestructive grade which, when exposed to pure oxygen, slowly acquires the explosive energy-density of dynamite. It should prove ideal for those fashionable architectural or technological monstrosities which, a few years later, turn out so horrific that they have to be destroyed.

David Jones

New influenza virus in horses

SIR — In March 1989, an outbreak of severe respiratory disease occurred in horses in the Jilin and Heilongjiang provinces of north-eastern China (Y. Guo *et al.* *Chinese J. exp. clin. Virol.* **3**, 318–322; 1990). Morbidity was 81 per cent and mortality as high as 20 per cent in some herds. In Zhilia county within Jilin province, 13,151 horses were infected and the mortality reached 35 per cent. Peaking in late March and April, the outbreak had abated by June 1989. Disease onset was acute, with coughing, bronchitis, pneumonia and fever of more than 40 °C. Animals were fatigued with inflammation of the nasal mucosa and copious mucous discharge from the eyes and nose. Most cases lasted for 5–6 days, but some animals showed symptoms for 2 weeks. Death was associated with pneumonia and enteritis. Similar signs of disease were noted in mules but not in donkeys. None of the people who worked with the diseased horses had signs of respiratory disease. A second outbreak occurred in April 1990 in Heilongjiang province, resulting in 41 per cent morbidity but no deaths.

An influenza virus of the H3N8 subtype was isolated from the infected animals and shown to be antigenically and molecularly distinct from the equine 2 (H3N8) and human H3N2 viruses currently circulating in the world. Sequence analysis of the entire haemagglutinin, nucleoprotein, matrix and nonstructural genes, together with partial sequencing of the three polymerase and the neuraminidase genes, established that five of the eight gene segments of this new virus, A/Equine/Jilin/1/89 (H3N8), are most closely related to avian influenza viruses of

Eurasian origin. The polymerase genes (PB1 and PB2) are genetically conserved between all influenza virus isolates and their origin could not be determined unequivocally. The partial sequence of the neuraminidase gene of the new H3N8 isolate was distinguishable from current equine H3N8 viruses, but limited data preclude establishment of its origin. (Y. Guo *et al.*, unpublished data.) Phylogenetic analysis indicated that four of the eight genes were recently derived from avian sources. The nucleoprotein gene was closely related to that found in the newly described H14 subtype isolated from ducks in the Soviet Union. On initial inoculation, the A/Equine/Jilin/1/89 (H3N8) influenza virus failed to replicate in ducks, but did replicate and cause disease in mice; on subsequent passaging it caused 100 per cent mortality. It caused severe influenza symptoms in ferrets.

The evidence suggests that this new equine influenza virus in northeast China is the latest mammalian influenza virus to emerge from the avian gene pool and that it has possibly spread to horses without gene reassortment. This direct transmission suggests that avian influenza viruses with this combination of surface proteins (H3N8) have a selective advantage over other subtypes for replication in horses. Seroarchaeological studies indicate that an H3N8 virus appeared in 1890 in the human population, coincident with historical records of an influenza pandemic in that year. Phylogenetic analysis of nucleoprotein genes suggests that the equine 2 virus genome originated in the late nineteenth century. Thus, one is tempted to speculate that the H3N8 human and equine viruses had similar origins.

The last known instance of an avian influenza virus transferring its entire genome to mammals and establishing a successful epidemic was in pigs in Europe (C. Scholtissek, H. Burger, P. A. Bachmann & C. Hannoun *Virology* **129**, 521–523; 1983). This avian-like swine influenza virus replaced classical swine influenza virus in pigs in Europe in the 1980s and has since become endemic in that swine population. There have probably been other successful transfers of intact avian influenza viruses to mammals. Available evidence suggests that the virus responsible for Spanish influenza in 1918–19 derived all its genes from an avian reservoir and entered the human population before 1918. The appearance of the new equine influenza virus in China emphasizes the possibility of whole avian influenza viruses entering mammalian hosts and serves as a warning that a new outbreak of human influenza could occur by this mechanism.

Will the outbreak in horses be limited to China, or could it become global? The answer will depend on the movement of horses from this region. Because there is trade across the Sino-Soviet border, the most

likely route for spread is through the Mongolian horse population to central Russia. People who keep horses should be aware of the presence of this virus and maintain quarantine practices to reduce its spread. One important consideration is that, although the virus originated recently from birds, it seems to have lost its ability to replicate in ducks, reducing the likelihood of spread by aquatic birds during migratory flights.

ROBERT G. WEBSTER

Department of Virology and
Molecular Biology,
St Jude Children's Research Hospital,
332 North Lauderdale,
PO Box 318, Memphis,
Tennessee 38101, USA

GUO YUANJI

National Influenza Centre,
Department of Influenza,
Institute of Virology,
Chinese Academy of Preventive
Medicine,
Beijing 100052,
China

Relationship of FKBP to PKC-2

SIR — Goebl *et al.* in *Cell*¹ and Tropschug and Hofmann in *Scientific Correspondence*² have reported the near-identity of the amino-acid sequence of the human FK506- and rapamycin-binding protein (FKBP)³ with a bovine protein, termed PKC-2 (ref. 4), which is purported to inhibit the enzymatic activity of protein kinase C (PKC). These authors suggested that PKC is the target of the immunophilin-drug complex². Indeed, the sequence reported for PKC-2 is identical to the amino-acid sequence of bovine FKBP purified from thymus⁵.

We have investigated the effect of FKBP on phosphorylation mediated by PKC and find that FKBP and its ligands FK506 and rapamycin, either acting alone or bound to FKBP, do not inhibit the kinase activity of isolated PKC or PKC-mediated events in cells⁶. Moreover, similar results have been obtained with cyclophilin and cyclosporin A (CsA). Thus, in the light of these biochemical and cellular studies, we surmise that the target protein(s) of immunophilin-drug complexes (FKBP-FK506, FKBP-rapamycin and cyclophilin-CsA) is not PKC.

MARK W. ALBERS

JUN LIU

STUART L. SCHREIBER

Department of Chemistry,
Harvard University,
Cambridge,
Massachusetts 02138, USA

Striding on water

SIR — Daedalus has suggested a 'Seasurfer sea-going platform' based on his recent invention, graphite-fluoride flock¹. Water striders (*Hemiptera: Gerridae*) have also exploited the hydrophobic properties of close-packed fibres to walk on water². Daedalus should consider the advantages inherent in constructing a platform based on water striders; the ability to move into the wind or up a wave slope to mention two. Furthermore, the ability of water striders to communicate by vibrating their forelegs on the water-surface membrane³ means that a Seasurfer constructed of water striders would come with its own built-in mobile phone.

STEPHEN J. CLARK

Field Research Center for Ecology and
Ethology,
The Rockefeller University,
Tyrrel Road, Millbrook,
New York 12545, USA

1. Jones, D. *Nature* **350**, 196 (1991).
2. Cheng, L. *Nature* **242**, 132–133 (1973).
3. Wilcox, R. S. *Science* **206**, 1325–1327 (1979).

1. Goebl, M. G. *et al.* *Cell* **64**, 1051–1052 (1991).
2. Tropschug, M. & Hofmann, R. *Nature* **351**, 195 (1991).
3. Standaert, R. F., Galat, A., Verdine, G. L. & Schreiber, S. L. *Nature* **346**, 671–674 (1990).
4. Mozier, N. M. *et al.* *Eur. J. Biochem.* **194**, 19–23 (1990).
5. Lane, W. S., Galat, A., Harding, M. W. & Schreiber, S. L. *J. prot. Chem.* **10**, 151–160 (1991).
6. Albers, M. W. *et al.* *BioMed. Chem. Lett.* **1**, 205–210 (1991).

Air pressure and methane fluxes

SIR — Mattson and Likens found¹ that sporadic methane bubble releases (ebullition) from lake sediments were correlated with changes in the local air pressure. A similar phenomenon has been known to colliery ventilation engineers in the United Kingdom for more than 250 years.

Saxton, in a series of articles² on the history of coal mine ventilation, has documented several accounts of investigations of the phenomenon of methane emission from underground cavities coincident with a falling barometric pressure. The first refers to investigations by Brownrigg prompted by the explosion of methane at the Corporal pit in Whitehaven in 1737. Brownrigg is said to have linked the possibility of such explosions with a rapidly falling barometer. The explosion at Haswell colliery, Durham in 1844 was investigated by Michael Faraday and Charles Lyell. The report suggested that a fall in the barometer resulted in the emission of methane into the airstream of the mine. Scott and Galloway³ investigated the statistical evidence for the linkage between methane explosions and the state of the weather, concluding that of the 525 explosions between 1868 and 1870, 49 per cent could be connected with disturbance of the barometer.

Following some disastrous explosions, a Royal Commission was set up in 1879 to inquire and report "whether, with respect to the influence of fluctuations in atmospheric pressure upon the issue of firedamp (methane) from coal, ... the resources of science furnish any practicable expedients that are not now in use and are calculated to prevent the occurrence of accidents, or limit their disastrous consequences". The Commission concluded that variations of atmospheric pressure exercise an undoubted effect on accumulations of gas in mines but noted widespread conflict of opinion as to the connection between atmospheric pressure and the occurrence of explosions.

The significance of changes in barometric pressure is recognized in the Acts and Regulations under which mines have to be operated. It is a statutory requirement that a barometer must be provided where it can be easily seen and read by people employed at the mine.

Even with modern ventilation standards, a rapid fall in barometric pressure can still cause serious problems. Hinchliffe⁴ describes the high gas emission at Allerton Bywater Colliery, Yorkshire in March 1986. This was the direct result of a severe drop in barometric pressure, which fell at approximately 3 millibars per hour for about 12 hours. High concentrations of methane were measured in the mine ventilation air and it was necessary to suspend some operations until the barometer rose again.

In all the above incidents, methane was

emitted from cavities of various kinds. Changes in barometric pressure will not affect the adsorption/desorption of gas from the coal itself, as the pressure associated with these processes are of the order of two or more atmospheres⁵. It is the time rate of change of air pressure that matters⁵. Mattson and Likens refer only to the changes in air pressure with no indication of the periods over which they took place: it would be interesting if they re-examined their data to see where the best correlation lies.

J. McQUAID
A. MERCER

*Health and Safety Executive,
Broad Lane,
Sheffield S3 7HQ, UK*

1. Mattson, M. D. & Likens, G. E. *Nature* **347**, 718-719 (1990).
2. Saxton, I. *The Mining Engineer* **145**, 496 & 556 (1985-86); **146**, 170 & 238 (1986-87).
3. Scott, R. H. & Galloway, W. *Proc. R. Soc.* **XX**, 292-305 (1872).
4. Hinchliffe, P. *The Mining Engineer* **146**, 479-484 (1986-87).
5. Carter, W. H. N. & Durst, C. S. *The Mining Engineer* **115**, 3-29 (1955).

In defence of Fourier

SIR — Maddox¹ and Malone and Wheatley² raise questions about the validity of Fourier's analysis of heat conduction and its equivalent for molecular diffusion. A particular point of dissatisfaction concerns the infinite speed of propagation of the diffusing property implied by the partial differential equation model.

We wish to draw attention to an early discussion of this problem by P. Frank³, who carefully points out that this diffusion theory is based on the assumption of a continuous distribution of matter in space and does not account for a finite number of discrete molecules. So we must interpret solutions to the diffusion equation in a probabilistic sense, representing the mean number of molecules in a certain region at a given time. Starting with a discrete number of molecules N at $x=0$ diffusing on the infinite x -axis, one can easily show that in time t , on average at most one particle will have diffused a distance at least

$$\xi = \xi(N, t) = 2f(N)\sqrt{Dt}$$

where D is the diffusion constant and $f(N)$ is the solution of

$$\psi(f(N)) = 1/N$$

here $\psi(x)$ denotes the complementary error function, and therefore $f(N)$ increases to infinity as N goes to infinity³. Thus, $\xi(N, \tau)/\tau$, where τ is an arbitrary unit of time, may be defined as the propagation speed of diffusion, infinite only in the limit as the number of molecules, N , goes to infinity. The concentration maximum in a variety of diffusion settings also occurs at distances of order $\sqrt{Dt}$ after t units of time⁴, a rule of thumb equally

suitable for defining a finite speed of propagation of diffusion³.

In view of this interpretation it seems to us that the current debate^{1,2} is about a rather small can of worms. Fourier's approach will retain its prominent place in our thinking of diffusion problems, and school teachers will wait for better reasons to change their curriculum.

VOLKER BRENDL
HAROLD LEVINE

*Department of Mathematics,
Stanford University,
Stanford, California 94305, USA*

1. Maddox, J. *Nature* **338**, 373 (1989).
2. Malone, P. C. & Wheatley, D. N. *Nature* **349**, 373 (1991).
3. Frank, P. *Physik. Zeitschr.* **19**, 516 (1918).
4. Crank, J. *The Mathematics of Diffusion* 2nd edn (Clarendon, Oxford, 1975).

A natural classification

SIR — Ernst Mayr criticizes¹ the new natural system we propose², and offers an alternative. Mayr has written of the need for consistency between phylogeny and taxonomy, and for precise measures of relationship among the higher taxa³, so we had anticipated his support.

Our system² is an attempt to bring taxonomic classification into line with the recent understanding of phylogeny that stems from molecular studies. It replaces the conventional 'default' system — an inconsistent amalgam of the five-kingdom classification of Whittaker and the modern version of Chatton's prokaryote-eukaryote dichotomy — which is phylogenetically incorrect. Mayr faults our system, however, for not reflecting the structural disparity between eukaryotes and prokaryotes — the latter having "relatively small differences between [them]"¹. Although he makes some minor terminological and relational rearrangements of categories, his alternative system is merely a return to what we consider to be the flawed conventional view that divides the living world into eukaryotes and prokaryotes.

Even if it is true, as Mayr says, that "the difference in structural organization between prokaryotes and eukaryotes is an order of magnitude greater than the relatively small difference between the Archaeobacteria and the Eubacteria"¹, this does not reflect an aboriginal phylogenetic division between prokaryotes and eukaryotes, nor does it mean that the archaea are related to the (eu)bacteria rather than the eukaryotes. Phylogenetically, the archaea's most recent common ancestor (determined by sequence comparisons using several unrelated pairs of paralogous protein genes) appears to have been with the eukaryotes, not with the bacteria^{4,5}. Thus, any system that divides the living world initially into prokaryotes and eukaryotes cannot be natural.

Many biologists tend to see morphological

complexity and diversity as being more significant than other kinds of complexity. But morphology is simply the most accessible aspect of organismic diversity because it is the most readily observed. In terms of their coding capacity, genetic organization and gene regulation, prokaryotes are no simpler than eukaryotes; and in terms of their metabolisms they are considerably more complex and diverse.

As Mayr cannot claim that his is a natural system, the difference between our systems may lie in a difference in our perceptions of the purpose of a taxonomy. He seems to suggest this when he justifies his system as one with which nonspecialists and "lay people... will be more comfortable"¹. In one sense, the taxa one defines are arbitrary constructs; they do not correspond to real (physical) entities. So why not define them in ways that feel familiar to us? But we must realize that the systems by which we organize our perceptions of the natural world do not merely reflect our views but also constrain them. The terms we use to describe systems and, even more, the formal structures of the systems we construct, determine the ways in which we can think about those systems. So it is important that we adopt terms and concepts that help rather than hinder our understanding. This consideration led us to propose the new 'domain' system and to replace the term 'archaebacteria' with 'archaea' to avoid further confusion between the two prokaryotic domains. And it is also this consideration that prevents us from returning to the artificial prokaryote-eukaryote dichotomy at the domain level.

CARL R. WOESE

Department of Microbiology,
University of Illinois,
Urbana, Illinois 61801, USA

OTTO KANDLER

Botanisches Institut,
Universität München,
8000 München 19,
Germany

MARK L. WHEELIS

Department of Microbiology,
University of California,
Davis,
California 95616, USA

1. Mayr, E. *Nature* **348**, 491 (1990).
2. Woese, C. R., Kandler, O. & Wheelis, M. L. *Proc. natn. Acad. Sci. U.S.A.* **87**, 4576-4579 (1990).
3. Mayr, E. & Ashlock, P. D. *Principles of Systematic Zoology* (McGraw Hill, New York, 1991).
4. Gogarten, J. P. et al. *Proc. natn. Acad. Sci. U.S.A.* **86**, 6661-6665 (1989).
5. Iwabe, N., Kuma, K., Hasegawa, M., Osawa, S. & Miyata, T. *Proc. natn. Acad. Sci. U.S.A.* **86**, 9355-9359 (1989).

Scientific Correspondence

Scientific Correspondence is intended to provide a forum in which readers may raise points of a scientific character. They need not arise out of anything published in *Nature*. In any case, priority will be given to letters of fewer than 500 words and five references. □

Morphologies of growth

SIR — Hill¹ proposes a criterion of selection for growth morphologies in which transitions between stationary states are "always into states of greater entropy production". This selection principle is *ad hoc* and does not hold for spatially homogeneous systems like the one that Hill considers.

Hill obtains the critical value of the 'driving force' X_c by equating the entropy production of two morphologies, which he claims is completely general as the linear phenomenological relation

$$V = L(X - \theta)$$

is "merely convenient and not essential". But he fails to realize that his condition, equation (4) will hold only if L_1 and L_2 are constant. For nonlinear phenomenological relations, $L = L(X)$, thus destroying his condition (4) in nonlinear regions. In fact, he claims that "in the linear regime these transitions cannot occur without discontinuities in the rate of crystallization because, as follows from equations (2) and (4), the only point free from such a discontinuity is the trivial one

$$(\theta_1 - \theta_2) \sqrt{(L_1 L_2)} = 0$$

in which the rates are zero". This means that the critical value, equation (4), of the driving force is valid only in the linear region, for when this value is introduced into the two phenomenological relations, one finds, on subtracting the two, that

$$V_1 - V_2 = (\theta_1 - \theta_2) \sqrt{(L_1 L_2)}$$

As the expression for the entropy production, Hill's equation (3), is no more general than the phenomenological relation, his equation (2), Hill's criterion in equation (4) is valid only for the linear regime. Hill has not shown that the intersection of the velocities determines the selection, as he has considered only the linear regime where there is always a discontinuity. Deviations from linearity in equation (2) would imply deviations from the parabolic form of the entropy, production (3), which would invalidate the conclusions obtained from Hill's Fig. 1.

The basic fallacy in Hill's argument is that he considers the system of "two morphologies formed by adjacent steady states" to be spatially homogeneous. I have shown²⁻⁴ that, in this case of spatial homogeneity, even with coupling, the system always tends to the state of lower entropy production as this lowers the velocity of the process. Hill's mechanism is equivalent to two uncoupled and spatially homogeneous chemical reactions with a common driving force. It is interesting that if the system is systematically displaced from equilibrium along the branch of solutions coinciding with the law of mass action at equilibrium, new branches can appear at a finite distance from equilibrium if the reactions are nonlinear. With the appearance of more than one branch, no branch can be

globally stable. Transitions to these kinetic (as opposed to the thermodynamic) branches will occur with a lower entropy production than the thermodynamic branch and, consequently, with a correspondingly lower velocity of reaction. This is what one would logically expect. The new kinetic branches ensure the stability of systems far from equilibrium.

This contradicts Hill's proposed criterion, according to which he claims, but does not demonstrate, the opposite. The situation in which the system is not spatially homogeneous is more complicated, but Hill does not consider this situation. He should, in fact, have done so as he is dealing with steady states with different spatial symmetries. But examples from chemical engineering⁵⁻⁷, such as the geometric instability in a porous catalyst particle, confirm the result that there is a transition to a steady state with lower entropy production and a corresponding lower velocity of reaction. Hence, Hill's selection principle is neither compelling nor convincing.

B. H. LAVENDA

Dipartimento di Scienze Chimiche,
Università degli Studi di Camerino,
Via S. Agostino,
1-62032 Camerino (MC), Italy

HILL REPLIES — The principle I suggest in my paper¹ is not confined to systems showing linear force-flow relations, nor is it applied to homogeneous systems.

Such a principle has no restriction to linearity but merely requires that the two entropy curves cross at a morphology transition. I chose to analyse a linear system, but entropy production curves in the region of the transition do not have to be parabolic and could have any form. Moreover, the slopes L and intercepts θ of linear equations may also be those of the tangents to two nonlinear force-flow curves in the region of the entropic crossover.

As for homogeneous systems, my paper¹ is about growth morphologies — it cannot be read in any way other than applying to dissipative structures. It certainly does not apply to "uncoupled and spatially homogeneous chemical reactions with a common driving force" because each morphology has its own driving force ($X - \theta$) but in a chemical system there would be only one force — the affinity A . My treatment is not confined to homogeneous systems and I cannot see any general proof in Lavenda's paper², which I am alleged to be contradicting, that such systems always tend to the state of lower entropy production. It is if true, it is interesting but irrelevant.

The crystal selection phenomenon I analysed is one in which entropy production is definitely increasing through the transition, and all the cases I am aware of in which different structures succeed each other in dissipative systems, from crystal forms, Bénard cells, Taylor vortices, Hele-Shaw growths to oscillatory patterns in the Belousov-Zhabo-

tinsky reaction, are all associated with increases in entropy production as the driving force is increased.

I suspect Lavenda feels I am violating some deep-seated principle that nature is always seeking stability either in equilibria or in minimum dissipation. I do not think such a principle exists, and if 'die Entropie der Welt strebt einem Maximum zu' then why should it not always strive to a maximum as fast as it possibly can?

ADRIAN HILL

Physiological Laboratory,
Downing Street,
Cambridge CB2 3EG, UK

- Hill, A. *Nature* **348**, 426–428 (1990).
- Lavenda, B. H. thesis University of Brussels, 1970).
- Glansdorff, P. & Prigogine, I. in *Thermodynamics of Structure, Stability and Fluctuations* Ch. 16 (Wiley-Interscience, New York, 1971).
- Lavenda, B. H. *Q. Rev. Biophys.* **5**, 429–479 (1973).
- van Heerden, C. *Ind. Engng Chem. (Prod. Res. Dev.)* **45**, 1242 (1953).
- van Heerden, C. *Chem. Engng Sci.* **8**, 133 (1958).
- Ans, R. & Amundsen, N. R. *Chem. Engng Sci.* **7**, 121; 132; 148 (1958).

Superantigen function

SIR — Superantigens combine with class II molecules of the major histocompatibility locus to form ligands capable of inducing powerful proliferative responses in CD4⁺ T cells that bear appropriate V β elements in their T-cell receptor. The prototype superantigens are encoded by the minor lymphocyte stimulatory (*Mls*) genes in mice and play an important role in selection of the T-cell repertoire during development by inducing clonal deletion of T cells expressing specific V β elements. No such elements have yet been identified in other species. Bacterial exotoxins are now recognized as acting as superantigens and this has stimulated further interest in the biological role of these proteins¹. The product of the open reading frame in the 3' long terminal repeat (LTR) of mouse mammary tumour virus (MTV) has recently been reported to encode the first identified viral superantigen². Here we report that herpesvirus saimiri (HVS), a herpesvirus of primates, encodes a homologue of the product of the 3' LTR open reading frame of MTV.

HVS establishes persistent asymptomatic infection of its natural host (*Saimiri sciureus*). But in other New World primates the virus induces rapidly progressive T-cell lymphomas. The pathology of these conditions varies between species, but is characterized by massive lymphocytic infiltration of many tissues. We have previously shown³ that an abundant transcript of 1.3 kilobase pairs (kbp) present at immediate-early times of infection with HVS (strain 11 (onc)) has the potential to encode a polypeptide of 249 amino acids, which is similar to the product of the conserved open reading frame in the 3' LTR of MTV, then of unknown function.

An alignment of the HVS protein with the

```

mtvc3h      MPRLOQKWLNSRECPTPRGEAAKGLFTTKDDPSAHKRMSPSDKDIFLLCKLGLTCLGLGEVVRARFATLD
qgmvm6m    MLLVLPRLOQKWLNSRECPTLRRAAKGLFTTKDDPSACTRMSPSDKDIFLLCKLGLTCLGLGEVVRARFATLD
HVS        MALLDRLMLKHLTINFSFRLMTI

mtvc3h      SFNSSSVQVYLNLSNSTFLLRQGPQPTSSYKPHRFQSEIEIRMLAKNYIFTNTPNIGRLVIMRNESLSFSNIFT
qgmvm6m    SFNSSSVQVYLNLSNSTFLLRQGPQPTSSYKPHRFQSEIEIRMLAKNYIFTNTPNIGRLVIMRNESLSFSNIFT
HVS        MLCIALFTSKPISITTEAPITNITQSPSLNISSSTLESEPLKNCCTFLDLWOLGENASIKDLMILQLR

mtvc3h      QIQLEPGIENKRKRSTSTEEVQGLLTTCIEVKKKKSEVFKIGDRWQPTSTYRGPYIYRPTIAPLPYTCGYDLNWD
qgmvm6m    QIQLEPGIENKRKRSTSTEEVQGLLRASQIEVKKRKRSTLVKIGDRWQPTSTYRGPYIYRPTIAPLPYTCGYDLNWD
HVS        DEVGSRITLPSPPSSQVEEQQLQRPNLPTAVGPPHKKYRLYNLWAPKADVNGKPIQFDLPPLPYTCGYDLNWD

mtvc3h      RWMIVNGYKVLRYSLPPRERILARPPWCNLSOBERDDMKQVHDYIYLGTMHFWGKIFIKKEGTVAGLTETHSYAKTY
qgmvm6m    RWMIVNGYKVLRYSLPPRERILARPPWCNLSOBERDDMKQVHDYIYLGTMHFWGKIFIKKEGTVAGLTETHSYAKTY
HVS        LMMNLNGKVRFDLSISWERTISGTPWCNTPSPSEAAALKQVILKAEKKRSTTSKSTIELEEQIKLEKETSPTSP

mtvc3h      GMSYEE
qgmvm6m    GMSYNG

```

Alignment of the sequences of the predicted protein product of the 1.3 kbp immediate-early transcript of HVS, the putative superantigen reported by Choi *et al.*² (mtvc3h) (ref. 2) and the product of the major open reading frame in the 3' LTR of an endogenous MTV (qgmvm6m) (ref. 3). Identical and similar residues are boxed. Similar residues were members of the sets: L=I=V=M, R=K=H, F=Y=W, S=T, E=Q, D=N. Amino acids 161–218 of the HVS protein, indicated by solid triangles, share an approximately 43 per cent identity and 60 per cent similarity with both murine proteins. The single, highly hydrophobic region in all three proteins is underlined.

putative superantigen reported by Choi *et al.*² and the homologous protein of a representative endogenous MTV is illustrated in the figure. Overall, the HVS and murine protein have an approximately 25 per cent identity and a 46 per cent similarity. Residues 161–218 of the HVS protein are 43 per cent identical and 60 per cent similar to both murine proteins and contain conserved acidic, basic and hydrophobic residues. In addition, all three proteins contain a single, highly hydrophobic region close to their predicted amino termini (see figure). Although correlations between the structure and function of viral superantigens have not yet been established, the similarities between the HVS and MTV proteins strongly suggest that they bear a significant evolutionary or functional relationship to each other. Superantigens encoded by bacteria do not share these structural features. No counterpart of this gene has been identified in other herpesviruses. HVS contains a number of genes that are likely to have been captured from the host cell (see refs 3 and 4). So the superantigen homologue could have been acquired from the natural host, implying the presence of *Mls*-like genes in the primate genome.

The mechanism by which HVS induces lymphoma is unknown, but is dependent on sequences from the 'left' end of the genome⁴. This region does not give rise to the 1.3-kbp transcript. In addition, most cell lines derived from HVS-induced lymphomas are of CD8⁺ phenotype. So there is no simple link between the superantigen homologue and oncogenicity. The gene may, however, play an important role in mediating the massive, early lymphoproliferation characteristic of HVS infection in primates other than the natural host. The administration of exogenous superantigen to immunologically

mature mice has recently been shown to induce a selective anergy⁵ of CD4⁺ cells bearing specific V β elements. Expression of a functional superantigen by HVS may therefore modulate the immune response to the virus and provide a novel mechanism of enhancing persistence in the natural host.

BRIAN J. THOMSON
JOHN NICHOLAS

Division of Virology,
National Institute for Medical
Research,
Mill Hill, London NW7 1AA, UK

Alzheimer's mutation

SIR — Goate *et al.* report¹ a mutation in the gene encoding the amyloid precursor protein in people with Alzheimer's disease in two unrelated families. The next step is to discover how commonly this mutation occurs in other cases of familial Alzheimer's disease.

We have screened an Italian and several French families that we have been studying^{2,3}. Although we do not find the mutation reported by Goate *et al.* in our Italian family, we do find that it occurs in three individuals in one of the French families⁴.

This result is particularly interesting in that one of these individuals has amyloid angiopathy. It is already known that a mutation reported by Levy *et al.*⁵ in the amyloid-precursor-protein gene is also the site of the mutation leading to hereditary cerebral haemorrhage with Dutch-type amyloidosis.

GERARD LUCOTTE
SALA BERRICHE
FABRICE DAVID

Regional Centre of Neurogenetics,
Robert Debré Hospital,
Rue Alexis Carrel, Reims, France

- Janeway, C. A. Jr *Cell* **63**, 659–661 (1990).
- Choi, Y., Kappler, J. W. & Marrack, P. *Nature* **350**, 203–207 (1991).
- Nicholas, J., Smith, E. P., Coles, L. & Honess, R. W. *Virology* **179**, 189–200 (1990).
- Biesinger, B., Trimble, J. J., Desrosiers, R. C. & Fleckenstein, B. *Virology* **176**, 505–514 (1990).
- Kawabe, K. & Ochi, A. *J. exp. Med.* **172**, 1065–1070 (1990).

- Goate, A. *et al.* *Nature* **349**, 704–706 (1991).
- David, F., Clerget, F. & Lucotte, G. *J. Neurol.* **235**, 485–486 (1988).
- David, F. & Lucotte, G. *Biomed. Pharmacother.* **42**, 575–578 (1988).
- Lucotte, G., David, F. & Hazout, S. *Ann. Génét.* (in the press).
- Levy, E. *et al.* *Science* **248**, 1124–1128 (1990).

A fox with quills

Robert W. Cahn

Sir Charles Frank, OBE, FRS: An Eightieth Birthday Tribute. Edited by R. G. Chambers, J. E. Enderby, A. Keller, A. R. Lang and J. W. Steeds. *Hilger: 1991. Pp.448. £27.50.*

MANY years ago, Isaiah Berlin, the historian of ideas, in seeking to come to grips with Leo Tolstoy's singular view of history, began by citing a line from an ancient Greek poet, Archilochus: "The fox knows many things, but the hedgehog knows one big thing." (Isaiah Berlin, *The Hedgehog and the Fox*; Weidenfeld and Nicolson, 1953.) "The hedgehog", Berlin wrote, "relates everything to a central vision", while "the fox pursues many ends, often unrelated and even contradictory." In analysing the mass of fruitful contradictions that was Tolstoy, Berlin proposed that "Tolstoy was by nature a fox, but believed in being a hedgehog." Sir Charles Frank began his professional life, perhaps, as a natural hedgehog, but his unquenchable curiosity turned him by degrees into what he is now, a fox furnished with quills.

Frank worked in the Physics Department at the University of Bristol from 1946 until his retirement in 1976, and has subsequently continued to play an active part in bristolian physics. A steady flow of wondrous physics has emerged from that department during the past half century, and Frank has been the fount, direct or indirect, of much of that. This volume incorporates 22 essays by eminent scientists in four continents — no fewer than ten of them fellows of the Royal Society — who either worked under his influence at Bristol or were elsewhere inspired by him during one of his many expeditions to overseas laboratories. The topics, taken in conjunction, eloquently illustrate a remark, cited here, which Frank made in his retirement speech in 1976: "Physics is not just concerning the nature of things, but concerning the interconnectedness of all the natures of things."

Scientific subjects covered by the essays include growth and morphology of crystals; interfacial structure, notably in epitaxial overgrowths; defect-induced pyroelectricity; liquid crystals; incommensurate and quasicrystals; stability of and defects in diamonds; polymer crystallization; creation of strong polymeric fibres by liquid- or solid-state processing; electrolytic 'parting' of gold alloys; the role of material flow in geophysics, especially in plate tectonics; several essays dealing with abstruse sidetracks of dislocation theory; and a remarkable one by L. I. Ponomarev of Moscow on muon-cata-

lysed nuclear fusion, a flourishing research field directly initiated by one of Frank's short but dense notes, dated 1947, itself occasioned by a mysterious observation by the Bristol cosmic-ray physicist, C. F. Powell, which caught Frank's roving attention, no doubt during laboratory tea. Indeed, there are many asides in this book about the fruitfulness of teatime conversations and I feel



Sir Charles Frank — "Like a chess virtuoso playing several games simultaneously . . .".

that a systematic study of the central role of laboratory tea, a largely British institution, in stimulating academic research is overdue. (I have much personal experience of the impact that this underrated institution has on foreign visitors to British universities!)

The most eloquent account of Frank in action at Bristol is to be found in the splendid essay by Andrew Keller on chain-folded crystallization of polymers. Keller depicts the Bristol laboratory as it first struck a brilliant, emotional, young Hungarian chemist: "Physics was in the air, was discussed everywhere: on the stairs, over tea, in the doors . . . passage of time was simply forgotten or ignored. There was no distinction between high and low brow, it was all an intellectual adventure. That is how polymers eventually

started in between quantum mechanics, particle physics, liquid helium, design of optical instruments, and much else. Here I saw science in action, not fragmented into specialities but as an indivisible whole, a single enterprise of the human mind . . . and Sir Charles was central to it all! Like a chess virtuoso playing several games simultaneously he was conducting these unforgettable teatime discussions on virtually all subjects in science."

The discovery of chain-folded crystallization of polymers, actually polyethylene, was a truly remarkable episode — wholly unexpected, *a priori* wildly improbable, as indeed were a number of other observations also recounted in the book which owed a good deal to Frank's stimulation. Keller's findings led to fierce and sustained controversy, only now gradually diminishing, as to the mode of formation of chain-folded crystals. His account, besides being a valuable historical record, also furnishes a sovereign contradiction of those deluded critics of the scientific method — we have all encountered them — who accuse its practitioners of an arid and passionless pursuit of bare, unaccommodated fact.

Some of Frank's ideas related to the work of nineteenth-century physicists, such as his very recent work on the vectorial depiction of the rotational relationships of neighbouring crystal grains (the Frank-Rodriguez method, here outlined by Robert Pond); he often found inspiration in physics of the distant past, in several languages, which is no doubt what incited the editors to include a stimulating essay by Frank Nabarro, full of nineteenth-century quotations in French and German, about a curiosity in crystal physics involving pyroelectricity in crystals which on the face of it should not exhibit it. Reverting to the Frank-Rodriguez approach, it is characteristic of Frank that the driving force for this work was his recollection of a practical problem put to him during a consulting visit to a US laboratory more than 30 years ago.

The first essay in the book, written by R. V. Jones with his inimitable brand of enthusiasm, describes Frank's student days and his subsequent work in intelligence during the war. In the episode of his successful interpretation of a mysterious air-reconnaissance photograph, his natural propensity to think in geometric terms is made evident. (A love of geometry is what made the early Frank a hedgehog.) This propensity is also consistent with his early concentration on dislocations, archgeometrical entities, which led in 1949 to his best-known prediction: that of dislocation-mediated crystal growth, unforgettably confirmed by an experimentalist minutes after the theoretical case has been publicly

expounded by Frank. (The prediction itself resulted from a mismatch between calculated and observed supersaturations for rapid crystal growth, an example of his habitual reliance on rigorous quantitative arguments.) The concept of the nonclosing Burgers circuit around a dislocation line motivated Michael Berry, in another splendid essay, into generalizing this notion into many other kinds of nonclosing circuits in several kinds of space, not necessarily 'real'.

From Berry's very abstruse piece of theory it is a long step to Kenneth Ashbee's account of bidimensional compression (originally

proposed by Frank with chain alignment of polymers in mind), now useful *inter multa alia* for extruding cider from apples, and to Donald Hurler's application of Frank's ideas of crystal morphology (elsewhere expounded in a lucid essay by P. Bennema) to the suppression of growth twins in semiconductor crystals. All is grist, or apples, that comes to the mill. □

Robert W. Cahn is in the Department of Materials Science and Metallurgy, University of Cambridge, Pembroke Street, Cambridge CB2 3QZ, UK.

Constructing 'everyanimal'

Mark Pagel

Body Size in Mammalian Paleobiology.

Edited by John Damuth and Bruce MacFadden. Cambridge University Press: 1990. Pp.397. £35, \$54.50.

E. O. WILSON once said of Carl Sagan "he wants to go to Mars. We haven't even landed on this planet yet. Let's put the money into research here." Palaeobiologists — people who reconstruct the biological past — often confront similar feelings from non-palaeobiologists: there is so much extant diversity to be explained that one wonders why palaeobiologists venture into the past. The present is rich in detail, the past hides all but a few fragments. Historical patterns must be inferred, current patterns can be measured and tested. There is always the worry that what one has constructed will come tumbling down with the unearthing of a new bone or dental fragment.

Palaeobiologists, not surprisingly, do not see it this way. To them, the distinction between the past and the present is artificial. Biological patterns exist in an expanse of time and space. Current patterns of biological diversity, they argue, are part of the same four-dimensional realm as historical evolutionary patterns: one cannot be understood if the other is absent. For example, we cannot fully appreciate the bones in our ears until we realize that at least one of them carries echoes of the ruminations of an ancient jawbone. More generally, there is the possibility of understanding who we are by first reconstructing who we were.

Here is a book about how to estimate the body size of extinct species from knowledge of how such things as teeth and bones — the things most likely to be preserved in the fossil record — relate to body size. Body size is the central player in palaeobiology because, as the editors say in the opening lines, "Body mass is correlated with a host of metabolic and physiological variables; with ecologically relevant characteristics such as life-history traits, diet, population density, population growth rate, home-range size, and beha-

vioural adaptations; and with large-scale patterns of community structure and biogeography". Such correlations are vital because none of these characters is preserved in the fossil record.

Palaeobiology can proceed in two directions. It can work forward from the past, or backwards from the present. Chroniclers of recent human social history — that is, people who write about Queen Bodicea and such — may use contemporary social theories to illuminate historical events, but never rely on the present to reconstruct the past. Increasingly, however, it seems that palaeobiologists use the present to reconstruct how the past must have been, in a way conceding that their historical records — the fossil beds — are depauperate. Palaeobiologists typically proceed by measuring in living species such things as teeth and bone dimensions and then estimating their relationships to body size. The equations derived from this exercise are then used to predict the body size of an extinct animal on the basis of, say, a fossil tooth. Once the species' body size is estimated, the fuller story can be written based on the empirical relationships of each of the many other characters — life histories and the like — also known to be related to body size. This places a premium on accurate estimation of the relationships of various traits to body size: a field known as allometry. In recognition of this, the editors encouraged their contributors to confront issues of measurement such as alternative indicators of prediction accuracy. Nearly every chapter does so; one is devoted to allometry, and there is an appendix of prediction equations contributed by the authors. This appendix, in combination with the chapters, makes the book a sort of do-it-yourself manual for reconstructing body size from teeth and bones.

Curiously absent from all of this attention, however, is discussion of two of perhaps the thorniest issues in allometry: what statistical model to use, and how to estimate the confidence one can place in the predictions derived from the model. Every analysis in this book uses standard regression estimates. Such estimates are known to underestimate the true value of the slope of the line relating, say, body mass to tooth size. Other techniques may not suffer as much from this problem. The contributors to this volume are

mute on this issue, apart from a passing remark in one chapter. Further, all analyses in this book are done using species as the data points. For example, one might examine the relationship of body size to femur length among living ungulate species. But statisticians and comparative biologists recognize that to treat species as independent data points yields artificially narrow confidence limits: we will unwittingly believe that our predictions are more accurate than they really are. None of the authors addresses this issue, and yet the reliability of the industry of constructing unknown bodies depends on it.

Examples of the unreliability of body-size estimates can be found among the contributions to this book: the *Australopithecus afarensis* female Lucy varies between 24 and 35 kilograms depending upon which of her bones is used to estimate her weight. In living mammals of the same body size, the proportions of soft tissue given over to bone, muscle and skin can vary considerably. Getting an estimate of body size can prove difficult even when the specimen is complete: one contributor to the book recounts how 76 individuals laboured to weigh a dead elephant. One cannot help but wonder how many of the three pages of the paper reporting their work are given over to listing the 76 authors' names.

But I think that the nearly obsessive concern with the techniques and pitfalls of predicting body size from fossil fragments reflected in this book can distract attention away from what may be a more central worry with the whole enterprise; a worry that I would like to be trivial but which may not be. In short, use of the present to reconstruct the past condemns the past to be like the present. Worse, perhaps, the past that we get from looking backwards is a very ordinary past, an average past. When we use a tooth to estimate an extinct mammal's body size, it is an average body size as derived from some equation. We will not know if the real thing had a larger or a smaller body for its tooth size than is expected from the equation. Then, when we substitute the estimated body size into an equation that predicts, say, the species' territory size from body size, we get a sort of second-order ordinariness. In short, the enterprise of deriving estimates based on estimates guarantees that the animals that get constructed are a kind of average animal — an 'everyanimal' of sorts.

Most of us have a fascination with human origins; something like the fascination with family genealogies but deeper, more primordial. This gives palaeobiologists, or as we used to know them, palaeontologists, a sort of hard-wired *raison d'être* in the public imagination. Damuth and MacFadden's book reveals how much wider is the scope of palaeobiology than just humans and human origins, or even just primates. It is a book with an honest approach, always vigilant that it is treading along a slippery path. Practitioners of palaeobiology will find it useful for having assembled a recondite set of facts and

for its deliberate cautiousness. But it seems to me that the central challenge to palaeobiologists must be to show that the patterns they infer are more than mere recapitulations of the present. This will never be easy: most of the inferences that palaeobiologists would like to draw will forever remain untestable because nature preserves very little other than the odd tooth or bone. There is reason for hope, though, I think. I just recently learned that my appetite for salted crisps may have its origin in the salt-poor diets of my hunter-gatherer progenitors. □

Mark Pagel is in the Department of Anthropology, Harvard University, Cambridge, Massachusetts 02138, USA.

Unanswered problems

John Brookfield

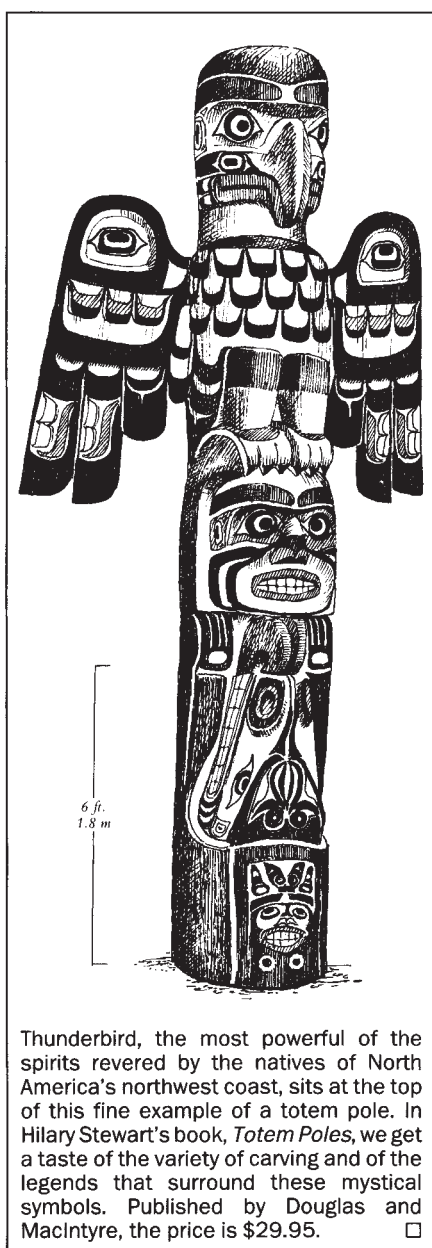
Fundamentals of Molecular Evolution. By Wen-Hsiung Li and Dan Graur. Freeman: 1991. Pp.284. £16.95, \$22.95 (pbk).

Evolution at the Molecular Level. By Robert Selander, Andrew Clark and Thomas Whittam. Freeman: 1990. Pp. 350. Hbk £44, \$55; pbk £23.95, \$28.95.

THE study of evolution has been transformed by rapid DNA sequencing methods, with the result that the vast new field of molecular evolution has arisen, whose dataset is dominated by the new DNA sequence arrays. These books differ in that Li and Graur present an undergraduate textbook, whereas Selander *et al.* have edited a collection of contributions from a symposium at University Park, Pennsylvania, in 1989, sponsored by the Society for the Study of Evolution.

The excellent introduction to molecular evolution by Li and Graur can be divided roughly into two parts. The first shows how sequences of the same gene from a variety of species are aligned, and then the DNA sequence information used, by distance or maximum-parsimony methods, to make an estimate of the true phylogenetic tree connecting the sequences. Given that the tree is known and dated by palaeontologists, one can arrive at estimates of evolutionary rates, in particular the synonymous and nonsynonymous rates of base substitution, K_s and K_a , in coding sequences.

I cannot recall a text in which the methods for these analyses are so clearly explained. DNA provides the best single dataset for the establishment of phylogeny. The dating of bifurcations using arguments based on the molecular clock has been less successful, due to changes in evolutionary rates, but remains the best method for groups with very poor fossil records. The main features of DNA



Thunderbird, the most powerful of the spirits revered by the natives of North America's northwest coast, sits at the top of this fine example of a totem pole. In Hilary Stewart's book, *Totem Poles*, we get a taste of the variety of carving and of the legends that surround these mystical symbols. Published by Douglas and MacIntyre, the price is \$29.95. □

sequence evolution are emerging, with K_a values being generally lower and more gene-dependent than K_s values, and evolutionary rates for introns, pseudogenes and 3' gene flanking sequences being approximately the same as K_s rates. The evolutionary conservation of important sequences is also a useful tool in the determination of function. What is more difficult is deciding what these results tell us about the types and proportions of substitutions that are selectively neutral. A strictly quantitative question about the neutral proportion of substitutions is of minor interest, because coding sequences form such a small proportion of eukaryotic DNAs that it is perfectly possible for only 0.5 per cent of base substitutions to be selected yet nevertheless for these to include the majority of nonsynonymous changes. One problem with molecular datasets is that there is no effective way to examine them that shows what proportion of amino-acid changes are selected. The obvious way is to look for

lineage and protein-dependent changes in substitution rates, suggesting effects of population size. Such changes are found, but their interpretation is complicated by the fact that the supposedly clock-like synonymous changes are also significantly variable between genes and lineages, perhaps because of variations in mutation rates.

The second part of Li and Graur's book is concerned with gene duplications and expansions and contractions in repeated-sequence families. This section is more descriptive, concerned mainly with listing the kinds of changes that have been seen to occur.

Evolution at the Molecular Level, which is of a consistently high standard, has a more prokaryotic flavour. Three elements can be distinguished as growth points, two of which relate specifically to bacteria, as discussed by Selander, R. F. Dubose, I. P. Crawford and others. First, for many organisms we have to reconsider the accepted view of allelic substitutions. In this view, new mutations arise which spread by drift or selection and replace the pre-existing allele. The assumptions are thus that there exists a population that is maintained at a stable size by ecological factors independently of the allelic substitutions taking place, and, second, that a new mutation can come by recombination to be in linkage equilibrium with the existing genetic variability, such that its spread is determined merely by sampling drift and its own selective effects. In sexual eukaryotes, these conditions may often be fulfilled. In bacteria, viruses and transposable elements, they certainly will not be. Bacteria are clonal and subject to cyclical selection. One consequence is that the effective population size is very much smaller than would be expected from the census size. Second, the use of DNA or any other characters to establish the phylogenies of organisms presupposes that there is a true phylogeny to establish. In bacteria, however, evolution can be highly reticulate. Finally, there has been the development of molecular population genetics, in which restriction-map or sequence information about variants within populations is obtained. The advance here is that DNA sequences may contain enough information to construct phylogenies of alleles. These may give evidence for balanced polymorphisms, as in the alcohol dehydrogenase gene of *Drosophila* (M. Kreitman) and mammalian MHC loci (M. Nei and P. Hedrick). They may also show ways in which gene trees differ from species trees, due to polymorphism at the time of speciation.

Both these volumes are highly recommended as giving a good insight into a field that, by happy coincidence, shares the possession of a set of very important unanswered problems and, now, the technical means for their solution. □

John Brookfield is in the Department of Genetics, University of Nottingham, Queen's Medical Centre, Nottingham NG7 2UH, UK.

Turning over pebbles

David Lindley

Lonely Hearts of the Cosmos. By Dennis Overbye. Macmillan: 1991. Pp. 438. £17.50, \$25.

OBSERVATIONAL cosmology is a fight over a factor of two. Hubble's constant, which relates a galaxy's redshift to its distance and therefore specifies how fast the Universe is expanding, is generally agreed to be no more than 100 kilometres per second per megaparsec, and no less than 50. In principle, figuring out where in that range it truly lies demands no more than the usual sort of stubbornness that any experimental science requires: galaxy redshifts are easy to measure, but estimating cosmological distances, by some independent means, is enormously difficult. Usually, the astronomer tries to find some proxy for the distance, such as the apparent brightness of an object whose intrinsic luminosity is known, but that brings in all kinds of complications. Is it safe to assume that the brightest hydrogen cloud in a distant galaxy or the brightest galaxy in a distant cluster are about as bright as the same objects in our own cosmic neighbourhood?

The search for the true value of Hubble's constant is not, therefore, a crazy race of the sort that particle physicists indulge in, when he who gets the most money and builds the biggest machine most quickly wins the prize, but a patient acquisition of data followed by minute efforts to tease out every possible source of bias and contamination. It attracts scientists of a certain mentality, people whose wish to know the design of nature is coupled with a willingness to turn over every pebble on the beach in the search. Such a person is Allan Sandage, the veteran astronomer who is the Levin (the blurb suggests King Lear, but that seems a bit too grand) of Dennis Overbye's saga of modern cosmology, *Lonely Hearts of the Cosmos*. Sandage evidently decided at an early age that his lot was to take over not merely the duties of his mentor, Edwin Hubble, but some version of his soul: in Sandage's world, there can be no cosmologist with more

burdensome a task, more seriousness of purpose, more freedom from frivolity, than Sandage.

All this would be wildly overblown if we were talking about a scientific effort to measure, for example, the energy-level structure of some complex atomic nucleus. Considered as a scientific job, measuring Hubble's constant is not significantly different from measuring anything else, but the fact that the outcome of the measurement tells us whether the Universe will expand for ever or one day collapse gives the research a certain weightiness, and makes it easier for the scientists involved to fall prey to arthurian delusions.

Most cosmologists, fortunately, stay relatively free of these over-romantic influences and, as Overbye ably recounts, spend their time as do modern scientists in general — holding raucous conferences in sunny locations (or wintry, for the skiing), and yelling at each other during seminars. Some readers of this book might find themselves put off by the author's frequent intrusions into the narrative, bumping about the Arizona desert in a jeep or frisbeeing on Hawaiian sands, but underneath this 'you-are-there' reportage, which seems to be a requirement for anything calling itself popular science these days, there is a shrewd and purposefully told tale. Few books mix the personal with the scientific, the narrative with the pedagogy, so tidily as this. And Overbye is by no means overawed by his subject — either the science or the people. The advantage of his spending some years as a sort of guest on the cosmological conference circuit is that he sees the characters not only as a collection of hard-working and creative scientists, but on occasions as quacks and peddlers of snake-oil. Astrophysicists, as Lev Landau is supposed to have said, are frequently wrong but never in doubt. Nor is Overbye overwhelmed by the wonders of modern cosmological theorizing — he seems in the end to side more with the observers, becoming dismayed and frustrated by the ease with which the theorists complicate the Universe a bit more to escape the latest observational constraint.

By the end of the tale, Sandage (who is the book's subject at least as much as modern cosmology) has been transformed by Overbye's close observation from a determined, if obsessive and over serious youth, into a much more ambiguous character. Ironically, for someone who set out to show that only by unceasing effort and painstaking work could the Universe be measured, Sandage in his later years seems to have followed many a great scientist in coming to believe that his experience and knowledge supply him with his own personal path to the truth, as if he had acquired a sixth sense by which he could distinguish good data from bad, and right answers from wrong. If Overbye wanted an epigraph for his splendid book, it could simply be: lighten up a little, Allan! □

David Lindley is an Associate Editor of Nature.

Solutions to solvation

Arthur Covington

Thermodynamics of Solvation: Solution and Dissolution. Ions and Solvents. Structure and Energetics. By G. A. Krestov. Horwood: 1991. Pp. 284. £48.

"SOLVATION should be considered as the sum of energetic and structural changes occurring in a system in the process of transferring gaseous ions (or other atomic-molecular particles) into liquid solvent, resulting in the formation of a homogeneous solution having a fixed chemical composition and structure," is the definition given by G. A. Krestov. He goes on to point out that the essential feature of solvation is a pictorial one of bonding of the solvent in the immediate vicinity of an ion forming a solvation shell. Problems arise in trying to determine the thickness of the shell and how many molecules it contains — the solvation number.

The information that thermodynamics can provide is limited by the fact that neutral combinations of ions have to be studied and that separating these into ionic contributions depends on arbitrary assumptions (conventions) or nonthermodynamic measurements; also, deriving structural information from these quantities is conjectural. Spectroscopic methods over the past 20 years have provided much more useful information on these points. Methods such as neutron and X-ray scattering and Raman and NMR spectroscopy are outside the author's remit and receive only brief mention. Neither do we find any details of the experimental methods used to provide the thermodynamic information. But there are many useful tabular data on the properties of ions and solvents as well as the expected free energies, enthalpies and entropies. There are also some unfamiliar quantities to intrigue — potential enthalpy, enthalpy capacity and isobaric-isothermic potential.

The translation editor, John Burgess of Leicester University, who has also written for the same publisher on this subject (*Metal Ions in Solution*, 1978), has done a good job on the Russian text, which dates from about 1982, although I detected some trivial errors — for example, in names (Hidman for Hindman, Solomon for Salomon). The expert will find in this book a good introduction to Russian work and thinking in this area; others will need to read Burgess's book first, or one of several alternatives, for the text is long on statements but short on explanations. I cannot agree with the publisher's flyleaf that this is a book for senior undergraduates or even for postgraduates. □

Arthur Covington is in the Department of Chemistry, Bedson Building, University of Newcastle, Newcastle upon Tyne NE1 7RU, UK.

New in paperback

- From Academic comes the fourth edition of *Physicochemical and Environmental Plant Physiology* by Park S. Nobel. Price is \$34.95 (also available in hardback, price \$69.95). □
- For an accessible reference to modern medical genetics see the revised and updated version of David Weatherill's *The New Genetics and Clinical Practice*. Publisher is Oxford University Press; price is £15. □
- Monoclonal antibodies, clonal selection and immunotoxins are just some of the topics covered in *Immunology: Recognition and Response* edited by William E. Paul. Publisher is Freeman; price is £10.95, \$12.95. □

Multiple intercellular signalling systems control the development of the *Caenorhabditis elegans* vulva

H. Robert Horvitz & Paul W. Sternberg

Developmental, genetic and molecular studies indicate that multiple intercellular signalling systems interact to specify the types and spatial patterns of cells generated during the formation of the vulva of the nematode *Caenorhabditis elegans*. Two classes of evolutionarily conserved transmembrane receptors and a Ras protein function in these signalling systems. The biology of vulval development provides a framework for understanding how cell interactions control the development of animals as diverse as nematodes, insects and mammals.

FORMATION of the vulva of the *Caenorhabditis elegans* hermaphrodite embodies many of the basic problems of developmental biology: the generation of cell diversity, cell growth and proliferation, the determination of cell type, cell differentiation, spatial patterning, cell-autonomous and cell-interactive events (classically referred to as mosaic and regulative development), morphogenesis, organogenesis, the control of developmental timing, and the generation of sexual dimorphism. Nonetheless, vulval development involves only a few cells. In addition, the cell divisions of vulval development occur over a period of only five hours, can be observed directly in living animals, and can be readily analysed using genetic and molecular methods. The *C. elegans* vulva is studied with the belief that general principles of development will emerge from a simple system that can be analysed easily and at high resolution.

Cell signalling is central to vulval development. The vulva is induced by the gonad from the epithelium that envelops the animal. Inductive signalling seems to function by preventing the action of another intercellular signal that otherwise inhibits vulval development. In addition, the responding vulval precursor cells interact to help to establish the precise spatial pattern of expression of vulval cell fates. A receptor tyrosine kinase, a Ras protein, and a member of the *lin-12/Notch* family of putative transmembrane receptors, act in the intercellular signalling systems that control vulval development.

We describe here the development of the *C. elegans* vulva, focusing on the role of cell signalling in this process and on how vulval development can inform us about aspects of animal development in general.

Summary of vulval development

The newly hatched larva of *C. elegans* is in many ways behaviourally and morphologically similar to the mature adult (see ref. 1 for a general review). Development after hatching involves the growth and proliferation of existing cell types and

the generation of new anatomical features, most of which are needed for reproduction. For example, the hermaphrodite, which is in essence a female that briefly produces sperm, generates structures used for copulation and for egg laying. One such structure is the vulva, which forms the external genitalia and connects the uterus with the outside environment (Fig. 1).

Six cells have the potential to generate vulval tissue²⁻⁵. These cells—P3.p, P4.p, P5.p, P6.p, P7.p and P8.p—are the posterior daughters of six of the twelve embryonically derived ectodermal P cells known as P1–P12 (Figs 2a, b, 3a)^{2,6}. (The cells and genes we discuss in this review are briefly described in Table 1.) The vulva derives from P5.p, P6.p and P7.p, which generate seven, eight and seven descendants, respectively (Figs 2c, 3b–f). By contrast, P3.p, P4.p and P8.p each generate two nonvulval descendants, which later fuse with an epithelial syncytium that envelops most of the animal and that is known as the hyp7 hypoderm. During the fourth larval stage, the 22 descendants of P5.p, P6.p and P7.p undergo a series of morphogenetic changes and cell fusions, resulting in a stack of six toroidal multinucleate cells, the centres of which constitute the vulval opening²⁻⁷. The cells of the vulva interact with specific nerve, muscle and uterine cells to form the multi-tissue structure used by the adult for copulation and egg laying^{5,7,8}.

Model for cell signalling

We first describe a model for the role of cell interactions and cell signalling in vulval development. We then summarize experiments that have led to this model and that have identified some of the genes and proteins that function in this process.

In the model (Fig. 4a), P3.p–P8.p can each express any of three distinct fates, that is, generate eight, seven or two descendants. These fates are called primary (1°), secondary (2°) and tertiary (3°), respectively. Cells that express the 1° and 2° fates generate distinct vulval cell types, whereas those that express the 3° fate generate a nonvulval cell type. The fate that each

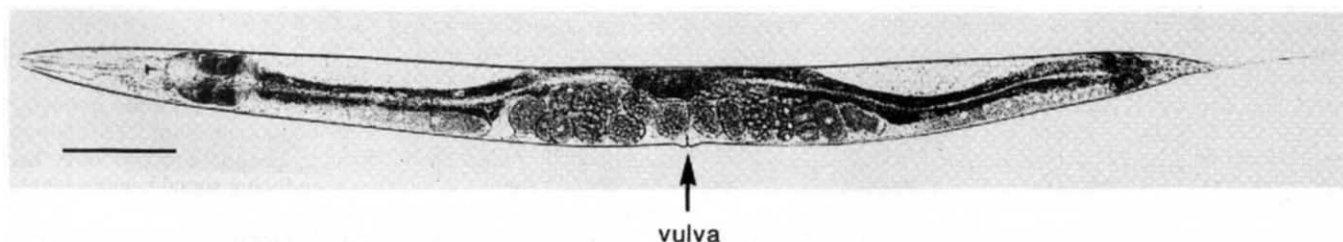


FIG. 1 Adult *C. elegans* hermaphrodite, bright-field optics. Lateral view—anterior, left; ventral, bottom. Arrow indicates vulva. Scale bar, 100 μ m.

TABLE 1 Key cells and genes

Cell(s)		
Pn.p	Set of 12 ventral epithelial cells; six can express vulval cell fates	
P6.p	Pn.p cell normally expressing 1° vulval fate	
P5.p, P7.p	Pn.p cells normally expressing 2° vulval fate	
P3.p, P4.p, P8.p	Pn.p cells normally expressing 3° nonvulval fate	
Anchor cell	Gonadal cell inducing vulval development	
hyp7	Syncytial hypoderm; might prevent vulval development	
Gene	Function in vulval development*	Protein product
<i>let-23</i>	Required for 1° and 2° fates	Receptor tyrosine kinase
<i>let-60</i>	Required for 1° and 2° fates	Ras protein
<i>lin-10</i>	Required for 1° and 2° fates	Novel
<i>lin-11</i>	Required for aspects of 2° fate	LIM homeodomain protein
<i>lin-12</i>	Required for 2° fate	<i>lin-12</i> /Notch-type receptor
<i>lin-15</i>	Required for 3° fate	Unknown

* Inferred from the phenotype of mutants in which gene activity is reduced or eliminated.

tripotent cell expresses is determined by cell interactions. First, a signal from the anchor cell of the gonad, which lies just dorsal to P6.p, induces vulval development by causing the three closest Pn.p cells to express the 1° or 2° fates. This signal seems to be both diffusible and graded, so that a high level results in the expression of the 1° fate and a low level results in the expression of the 2° fate. Without the signal, only the 3° fate is expressed (Fig. 4b). Induction might act by overcoming an inhibitory effect of the hyp7 syncytial hypoderm on vulval development, because it seems that elimination of a signal from the hyp7 syncytial hypoderm causes all six of the Pn.p cells to express vulval fates (Fig. 4c). Finally, a Pn.p cell that expresses the 1° fate can cause its Pn.p cell neighbours to express the 2° instead of the 1° fate. Without this lateral signal, which could act through direct cell-cell contact, neighbouring Pn.p cells that have been induced by the anchor cell can express the 1° fate (Fig. 4d). The inductive and lateral signalling processes might be partially redundant, so that lateral signalling reinforces the 2°-1°-2° spatial pattern of Pn.p cell fates established by a graded inductive signal.

The three developmental potentials

The latent developmental potentials of the tripotent Pn.p cells have been revealed by perturbing normal development. First, after destruction of one Pn.p cell by laser microsurgery, a nearby Pn.p cell can replace it and express its fate^{3,4}. For example, when P6.p is killed, P5.p can replace it and express the 1° fate instead of the 2° fate. Thus, P5.p is at least bipotent, and the presence of P6.p normally prevents P5.p from expressing the 1° fate rather than the 2° fate. Second, in mutants for either the *lon-1* or *dig-1* gene (for 'long' and 'displaced gonad', respectively), the gonadal anchor cell is displaced with respect to the Pn.p cells, and the pattern of expression of Pn.p cell fates is altered^{4,5}. Such studies established that all six cells have the potential to express any of the three P3.p-P8.p cell fates.

Because of their equivalence in developmental potential, P3.p-P8.p are said to constitute the vulval 'equivalence group'^{3,4}. Cell replacements among P3.p-P8.p occur in a hierarchy: in wild-type animals, P6.p can be replaced by any of the other five cells; P5.p and P7.p can be replaced by any of the cells except P6.p; and P3.p, P4.p and P8.p are not replaced. It is for this reason that the eight-, seven- and two-descendant fates are known as 1°, 2° and 3°, respectively. These fates are normally expressed in an anteroposterior pattern of 3°-3°-2°-1°-2°-3°. Thus, cell interactions specify the spatial patterning of cell fates during vulval development.

Similar experiments suggest that, unlike Pn.p cells, the daughters of Pn.p cells are not equivalent in their developmental potentials⁴. For example, after ablation of some of these daugh-

ter cells, the others do not change their fates. Thus, it seems that once the fate of the Pn.p cells has been determined, these cells divide to generate descendants of different types through mechanisms that probably do not involve further cell interactions.

A diffusible and spatially graded inductive signal

Vulval development depends upon an inductive signal from the nearby gonad. If the gonad is destroyed using a laser microbeam, no vulva is formed³. The gonadal signal emanates from the anchor cell (see Fig. 3a): if the anchor cell is destroyed, P3.p-P8.p all express the 3° nonvulval fate, and the resulting animal cannot copulate or lay eggs; furthermore, the anchor cell alone is sufficient among gonadal cells to induce vulval development⁹. Thus, the anchor cell induces three of the six Pn.p cells of the vulval equivalence group to undergo vulval development. The anchor-cell signal apparently can act at a distance, because in some *dig-1* mutant animals the gonad is displaced to the dorsal side, but nonetheless induces vulval development by the Pn.p cells, which are located along the ventral side⁵. Thus, the inducing signal probably either is diffusible or can be transmitted through other cells of the animal.

How does the anchor cell induce two distinct fates in the specific spatial pattern 2°-1°-2°? The inductive signal could be permissive, allowing three Pn.p cells to express either the 1° or 2° fate but not specifying which fate each cell will express; in this case the 2°-1°-2° pattern of cell fates could depend either on intrinsic differences among the Pn.p cells (pre-patterning) or on interactions among the Pn.p cells (self-organization by lateral signalling). Alternatively, the inductive signal could induce only the 1° fate, and a Pn.p cell so induced could cause its neighbouring Pn.p cells to express the 2° fate (sequential inductions, with inductive signalling followed by lateral signalling). A third possibility is that the inductive signal is instructive; Pn.p cells closer to the anchor cell could express the 1° fate because they receive a signal that is stronger or earlier than, or molecularly distinct from, that received by Pn.p cells that are further from the anchor cell and express the 2° fate.

Proximity to the anchor cell defines which Pn.p cell will express the 1° fate, even in *lon-1* and *dig-1* mutants in which the anchor cell and the Pn.p cells are displaced relative to one another^{4,5}. Thus, the expression of the 1° and 2° fates does not seem to be pre-patterned. Although some degree of intrinsic bias could exist among P3.p-P8.p (for example, P6.p could be tripotent but still have a tendency to express the 1° fate), all observations are consistent with these six cells being identical in their developmental potentials.

Several studies indicate that the inductive signal can specify both the 1° and 2° fates directly, without requiring interactions among the Pn.p cells. Specifically, in animals with only one Pn.p cell (either laser-operated wild-type animals or *unc-84* mutant animals) that cell can express either the 1° fate or the 2° fate⁴ (P.W.S., M. Herman and H.R.H., unpublished observations); a single Pn.p cell close to the anchor cell expresses the 1° fate, whereas one farther away expresses the 2° fate. In addition, Pn.p cells expressing the 2° fate have been observed in *dig-1* mutant animals in which no Pn.p cell expressed the 1° fate⁵. These findings suggest that lateral signalling among the Pn.p cells is not necessary for expression of either the 1° or 2° fate. However, these experiments could be misleading. For example, Pn.p cell debris in a laser-operated wild-type animal could act like an intact Pn.p cell.

Current data do not distinguish between a model involving a single graded anchor cell signal with two thresholds of response (a higher threshold for 1° than for 2°), and a model involving two distinct signals in which a 1°-inducing signal is more highly localized than a 2°-inducing signal.

As we discuss below, other studies indicate that lateral signalling among the Pn.p cells is also involved in determining the spatial pattern of Pn.p cell fates.

An inhibitory signal from the hypoderm

Activity of the gene *lin-15* (for 'cell lineage abnormal') inhibits the expression of vulval cell fates. When *lin-15* activity is reduced by mutation, animals form multiple vulva-like structures as a result of all six (rather than only three) of the cells P3.p–P8.p expressing either 1° or 2° vulval fates^{10,11}. Destroying the gonad in *lin-15* animals does not prevent the expression of this 'multi-vulva' phenotype, indicating that the *lin-15* mutation does not act by causing the overproduction of inducing signal within the gonad¹¹.

Genetic mosaic analyses of *lin-15* confirmed that the *lin-15* mutation acts outside the gonad and indicated that it acts outside the P3.p–P8.p cells¹². In these studies, animals having both genetically mutant and wild-type cells were generated and examined to determine which cells must be genetically mutant for the mutant phenotype to be expressed¹³. This approach reveals whether a gene acts within the cells it affects or whether it acts within other cells that influence the affected cells through cell–cell interactions. P3.p, P4.p or P8.p cells that are genetically wild-type for *lin-15* were found to be able to express the 1° or 2° vulval fates if cells other than the P3.p–P8.p cells are genetically mutant for *lin-15*. Thus the *lin-15* mutation acts elsewhere than in P3.p–P8.p.

Together, these laser-ablation experiments and mosaic analyses suggest that the *lin-15* gene functions in some cell(s) other than the anchor cell and the Pn.p cells. Herman and Hedgecock¹² postulated that the cell in which *lin-15* functions is *hyp7*, the syncytial hypodermal cell that envelops most of the animal. Specifically, they proposed that *hyp7* normally prevents Pn.p cells from expressing the 1° and 2° fates, and that *lin-15* activity is necessary for this inhibition. A mutation that prevents *lin-15* from functioning therefore causes all Pn.p cells to express either the 1° or 2° fate. This hypothesis suggests that induction by the anchor cell acts to override an inhibitory influence by *hyp7* and thus allow P5.p–P7.p to express vulval fates instead of the 3° nonvulval fate. Because the 3° fate involves cell fusion between *hyp7* and the Pn.p cells, *hyp7* could inhibit expression of the

1° and 2° vulval cell fates simply by fusing with Pn.p cells. If so, the inductive signal would act to prevent this fusion.

The anchor cell could antagonize an inhibitory signal from *hyp7* by acting on either *hyp7* or the Pn.p cells (or both). Because the anchor cell affects Pn.p cell fates in *lin-15* mutants^{14,15}, which presumably lack the inhibitory *hyp7* influence, the simplest hypothesis is that the anchor cell acts directly on the Pn.p cells.

The involvement of *hyp7* in vulval development elegantly explains the results of *lin-15* mosaic analyses. For this reason, we include *hyp7* in our model. However, the evidence for *hyp7* involvement is circumstantial. Furthermore, a *lin-15* mutation could result in an abnormal expression of the anchor cell signal (for example, by *hyp7*), causing gonad-independent vulval development, even though normally only the anchor cell and the Pn.p cells are involved in vulval development.

Many genes function in inductive signalling

Hundreds of mutants defining more than 40 genes involved in vulval development have been identified and characterized^{10,11,15–21} (S. Clark, Jeffrey Thomas, S. Kim, James Thomas and R. Horvitz, unpublished results; G. Jongeward, M. Han, J. Lee, R. Hill and P. Sternberg, unpublished results). Some of these mutants lack a vulva, and are said to be vulvaless (Vul). Because *C. elegans* hermaphrodites are internally self-fertilizing, Vul animals produce fertilized eggs that cannot be laid; these eggs develop *in utero* and hatch into larvae that crawl free from their parent. Other vulval mutants (such as *lin-15* mutants) have multiple vulva-like structures and are called multivulva (Muv).

We focus here on the Vul and Muv mutants that are perturbed in the cell interactions involved in vulval development. Mutants with altered patterns of expression of the 1°, 2° and 3° fates seem likely to define genes that function in cell signalling, because the pattern of Pn.p cell fates is determined by cell interactions. Such mutants define at least 30 genes. In Vul mutants altered in patterns of P3.p–P8.p cell fates, none of the Pn.p cells expresses the 1° or 2° fate; instead all six express the 3° fate. These mutants look like animals in which the anchor

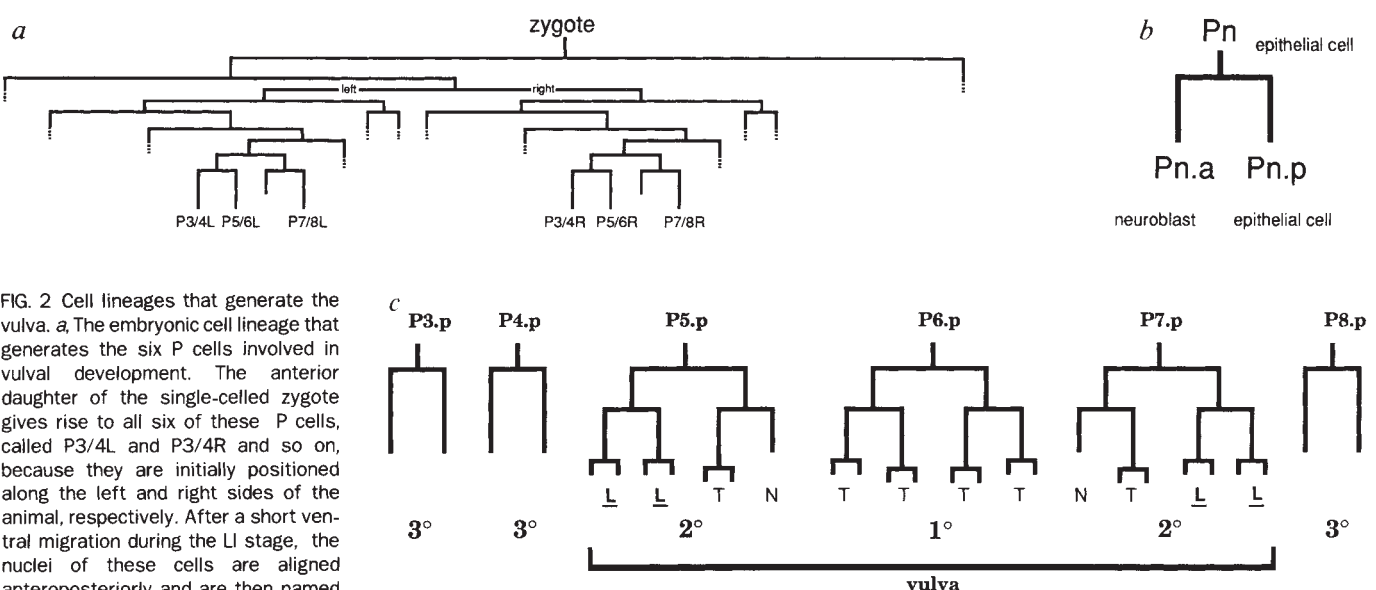


FIG. 2 Cell lineages that generate the vulva. *a*, The embryonic cell lineage that generates the six P cells involved in vulval development. The anterior daughter of the single-celled zygote gives rise to all six of these P cells, called P3/4L and P3/4R and so on, because they are initially positioned along the left and right sides of the animal, respectively. After a short ventral migration during the L1 stage, the nuclei of these cells are aligned anteroposteriorly and are then named P3–P8 based on their positions. The dashed lines indicate cells that divide embryonically. All divisions shown, except for the single left–right division indicated, occur along the anteroposterior axis. Anterior daughters are depicted to the left on the diagram. Adapted from ref. 6. *b*, Each epithelial Pn cell ($n=3-8$ for the cells involved in vulval development) divides during the L1 stage to generate an anterior neuroblast Pn.a and a posterior epithelial cell Pn.p. *c*, Vulval cell lineages. Each P3.p–P8.p cell divides anteroposteriorly according to the pattern shown, with each anterior daughter depicted to the left on the

diagram. The eight-descendant fate normally expressed by P6.p is called 1°; the seven-descendant fate normally expressed by P5.p and P7.p is called 2°; and the two-descendant fate normally expressed by P3.p, P4.p and P8.p is called 3° (see text). The vulval cell types are defined by two criteria: the axis of the third round of division (L, longitudinal; T, transverse; N, no division), and adherence to the ventral cuticle, indicated by **L**. There are also ultrastructural differences among these cell types⁵⁰ (J. G. White, personal communication). Adapted from ref. 4.

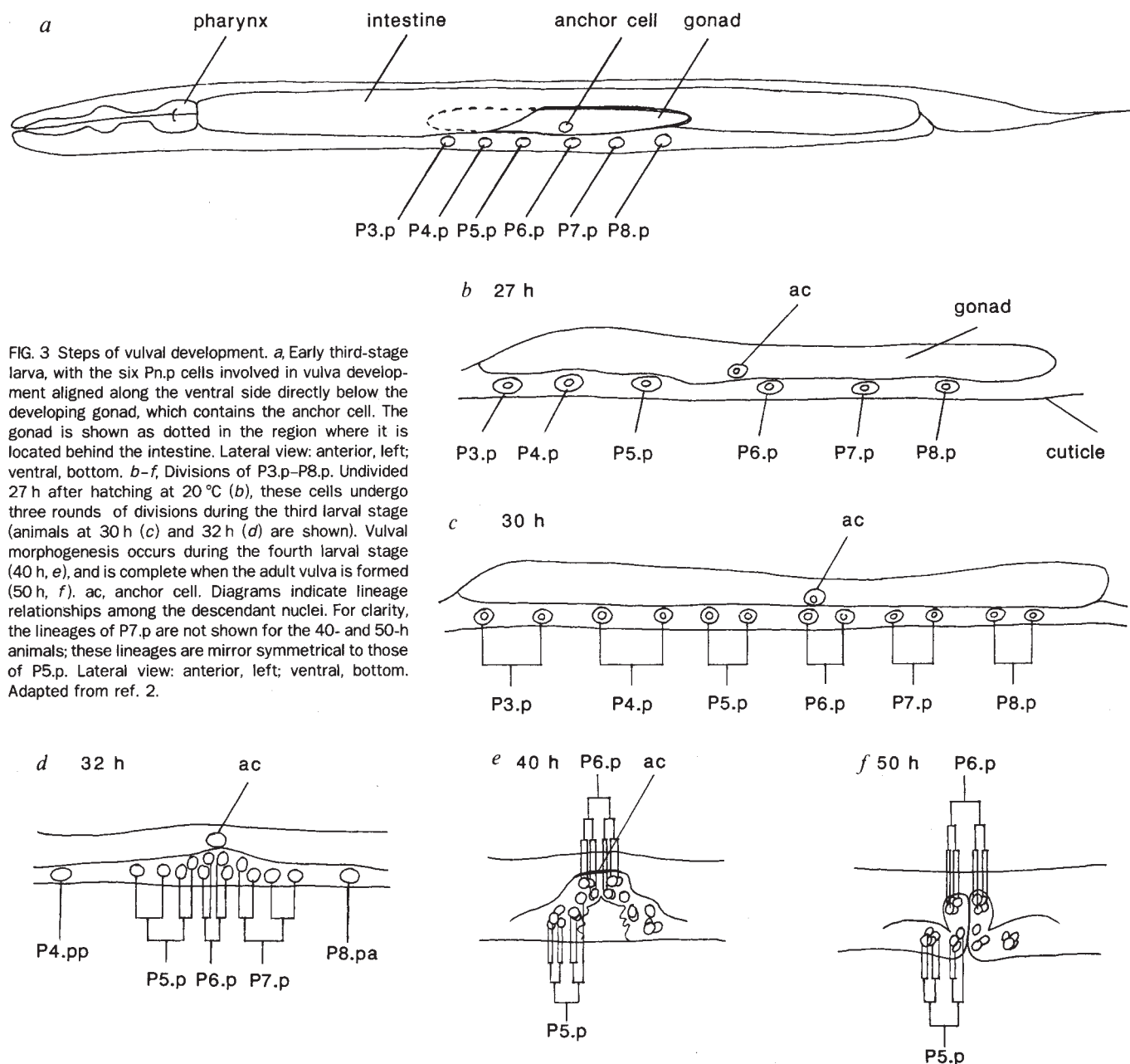
cell has been destroyed and in general probably define genes that are necessary for inductive signalling. In Muv mutants altered in patterns of P3.p–P8.p cell fates, all of the six P3.p–P8.p cells express the 1° or 2° fate. These Muv mutants in general define genes that prevent the expression of the vulval fates. Because these mutants are Muv even if the gonad has been destroyed with a laser microbeam, the mutations responsible can act outside the gonad. Determining whether the Vul and Muv mutations act in the gonad, *hyp7*, the Pn.p cells or elsewhere is crucial for understanding how the corresponding wild-type genes function.

One gene involved in inductive signalling is known to act as a binary switch in this process. Mutations that decrease *let-60* (for 'lethal', as some alleles result in lethality) function cause a Vul phenotype, whereas those that increase *let-60* function cause a Muv phenotype, even in the absence of the inductive signal^{15,21}. Inappropriate expression of the wild-type *let-60* gene also causes a Muv phenotype²². Thus, the state of *let-60* activity can determine Pn.p cell fates: if *let-60* is inactive, P3.p–P8.p cells express the nonvulval 3° fate; if *let-60* is active, P3.p–P8.p

cells express the vulval 1° or 2° fates. These findings strongly suggest that the inductive signal activates *let-60*.

On the basis of its DNA sequence, *let-60* encodes a protein that is 83% identical in the first 164 of its 184 amino acids to the human N-Ras protein²². Ras proteins bind GTP and GDP and are thought to act as GTP-activated switches in signal transduction pathways in both yeast and mammalian cells^{23–26}. That *let-60* encodes a Ras protein both suggests biochemical aspects of the role of *let-60* in vulval induction and reveals that a *ras* gene is essential for cell signalling in normal animal development. Mutations that activate *ras* in mammalian cells are oncogenic and constitute one of the most frequent types of mutation associated with human cancers²³. Muv mutations that activate *let-60* alter codon 13 (ref. 21), which is also mutated in certain oncogenic *ras* genes in mammals²³.

Another gene defined by Vul mutants and necessary for vulval induction also encodes a protein similar to proteins involved in cell signalling: the presumptive product of the *let-23* gene is a member of the epidermal growth factor (EGF)-receptor tyrosine kinase family²⁷. Receptor tyrosine kinases act by binding ligand,



transducing a signal across the plasma membrane, and catalysing the phosphorylation of specific tyrosine residues of cytoplasmic substrate proteins²⁸. In addition, a member of the EGF receptor tyrosine kinase family seems to be the receptor for intercellular signals during the development of *Drosophila melanogaster*²⁹⁻³². Thus, *let-23* is an excellent candidate for encoding the receptor for the inductive signal. Increased activity and increased dosage of the *let-60* gene bypass the effects of reduced *let-23* function²². This result would be expected if *let-60* acts in response to *let-23*. One plausible model is that the *let-23* receptor tyrosine kinase is localized to the membrane of the Pn.p cells and responds to inductive signalling by activating the *let-60* Ras protein in the Pn.p cells. However, the site(s) of action of *let-23* and *let-60* are unknown, and other models for how these genes function remain possible.

A third gene defined by Vul mutants is *lin-10* (ref. 20). The amino-acid sequence of the putative protein product of *lin-10* is not similar to that of any other known protein.

Lateral signalling and the 2° fate

The specification of the fates of the Pn.p cells not only involves signalling from the gonad and, perhaps, from the hyp7 hypoderm, but also involves signalling among the Pn.p cells themselves. In Muv mutant animals that lack an anchor cell,

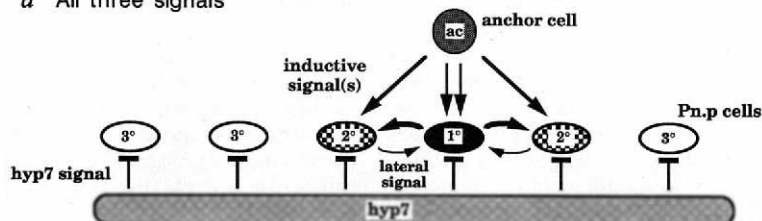
Pn.p cells generally express the vulval 1° and 2° fates in an alternating pattern, for example 2°-1°-2°-1°-2°-1°^{14,33}. Exceptions to this pattern are almost always adjacent pairs of 2° cells; adjacent pairs of 1° cells are not observed. Similar alternating patterns of Pn.p cell fates and an absence of pairs of adjacent 1° cells are seen in *dig-1* mutants⁵. Thus, expression of the 1° fate could cause neighbouring Pn.p cells to express the 2° fate and prohibit them from expressing the 1° fate. Support for this hypothesis comes from studies of P7.p in *lin-15* Muv animals¹⁴. If the anchor cell and all five cells of the vulval equivalence group except P7.p are killed, P7.p invariably expresses the 1° fate. By contrast, if P8.p is also present, P8.p sometimes expresses the 1° fate, in which case P7.p can express the 2° fate. Therefore, the fate of P7.p can be influenced by P8.p, establishing that Pn.p cells can communicate with each other. But this lateral signalling has been directly demonstrated only in *lin-15* mutants and has not been shown to occur during wild-type vulval development.

The lateral signal is probably mediated by the product of the gene *lin-12*: in a *lin-15* double mutant with a *lin-12* loss-of-function mutation, adjacent Pn.p cells can express 1° fates³³. In addition, expression of the 2° fate, which can be caused by lateral signalling, depends on *lin-12* function in non-*lin-15* animals: in animals in which *lin-12* function is eliminated, no

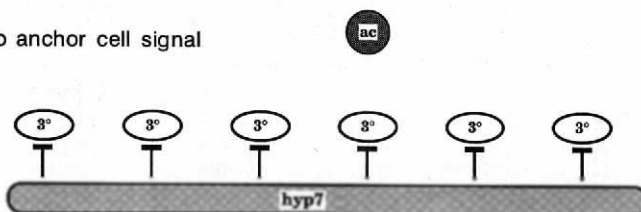
FIG. 4 Models for the roles of cell signalling in vulval development.

a, Overall model. Three intercellular signals control the expression of vulval cell fates. See text for general description. Inductive signalling from the anchor cell might involve either one or two molecularly distinct signals⁴. In a one-signal model, the anchor cell emits a spatially (or temporally) graded signal; a high level (or early reception) of the signal induces the 1° fate, a lower level (or later reception) induces the 2° fate, and no signal or a level of signal that is below a threshold fails to induce vulval cell lineages and results in expression of the 3° fate. In a two-signal model, the anchor cell emits one signal that has a highly localized distribution and a second signal that is more broadly distributed; the first signal (possibly in conjunction with the second signal) induces the 1° fate, and the second signal induces the 2° fate. The anchor-cell inductive signal might activate the lateral signal (see text), and so we show the lateral signalling process as involving only the Pn.p cells that express the 1° and 2° fates; although the lateral signal from the 2° Pn.p cells might be bidirectional, because this signal does not seem to affect the 3° Pn.p cells, we do not show a 2°-3° lateral signal. **b**, In the absence of the anchor-cell signal, the hyp7 inhibitory signal prevents all six Pn.p cells from expressing the 1° or 2° fates. Again, no lateral signal is shown. The consequence of eliminating the anchor-cell signal is known from animals in which either the gonad primordium or the anchor cell was destroyed using a laser microbeam^{3,9}. **c**, In the absence of inhibitory signal from the hyp7 syncytial hypoderm, each of the six Pn.p cells expresses either the 1° or the 2° fate. Lateral signalling causes the neighbours of 1° Pn.p cells to express the 2° fate, and the anchor cell causes the nearest Pn.p cell to express the 1° fate. The pattern of 1° and 2° cell fates expressed by the cells farthest from the anchor cell (P3.p, P4.p and P8.p) varies among individuals. The consequence of eliminating the hyp7 signal is inferred from the phenotype of *lin-15* mutant animals, which presumably lack this signal¹². **d**, In the absence of the lateral signal, the anchor-cell signal induces nearby Pn.p cells to express vulval cell fates and the Pn.p cell neighbours of 1° cells can express the 1° fate because they are not made to express the 2° fate by lateral signalling. The consequence of eliminating the lateral signal is inferred from the phenotype of *lin-12* loss-of-function mutant animals, which probably lack the response to this signal³³. As discussed later in the text, if *lin-12* functions in a 2°-specific inductive signalling pathway as well as in the lateral signalling pathway, *lin-12* loss-of-function mutants would be defective in two signalling pathways; in this case, specific elimination of the lateral signalling pathway could result in a normal 2°-1°-2° pattern rather than an abnormal 1°-1°-1° pattern of Pn.p cell fates.

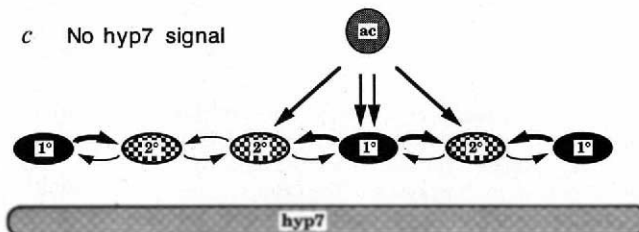
a All three signals



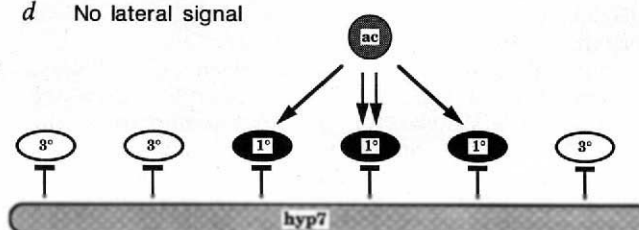
b No anchor cell signal



c No hyp7 signal



d No lateral signal



Pn.p cell expresses the 2° fate; in animals in which *lin-12* function is increased, all six Pn.p cells can express the 2° fate¹⁸.

Genetic mosaic analyses of another cell-fate decision controlled by *lin-12* demonstrates that *lin-12* acts in responding cells rather than in signalling cells³⁴. On the basis of its DNA sequence^{35,36}, *lin-12* is predicted to encode a transmembrane protein with 13 copies of an EGF-like motif, three copies of a different cysteine-rich motif, and six copies of the cdc10/SWI6 repeat, which is found in the protein products of a variety of genes^{37,38}. These features define the *lin-12* protein as a member of a family conserved in nematodes, insects and vertebrates³⁹⁻⁴¹. One member of this family, the *D. melanogaster* Notch protein, is associated with the plasma membrane^{42,43}. Together these observations strongly suggest that the *lin-12* protein acts as a receptor in vulval development.

From genetic and molecular analyses of *lin-12* mutants, Greenwald and Seydoux⁴⁴ proposed that mutations that cause increased *lin-12* activity result in ligand-independent activation. The ligand for the presumptive *lin-12*-encoded receptor is unknown. The *Drosophila* Notch protein seems likely to bind molecules located on the surface of neighbouring cells^{42,43}. Greenspan⁴⁰ has suggested that the Notch protein is a cell adhesion molecule that conveys information from the outside of the cell to its cytoplasmic cytoskeleton, with subsequent changes in cell shape altering patterns of gene expression. The *lin-12* protein might function similarly, acting by direct cell-cell contact and mediating changes in cell adhesion between adjacent Pn.p cells.

Because Pn.p cells can express the 2° fate in the absence of other Pn.p cells, as described above, it seems likely that lateral signalling between Pn.p cells is not essential for induction of the 2° fate. Thus, *lin-12* might be activated by inductive signalling from the gonad, either by a distinct 2°-specific inductive signal³⁹ or cell-autonomously (perhaps via an autocrine process) by a low level of a single inductive signal³³.

Inductive and lateral signalling interact

Inductive signalling causes Pn.p cells close to the anchor cell to express the 1° or 2° fate, whereas Pn.p cells further from the anchor cell express the 3° fate. Either a graded inductive signal or a lateral signal could be responsible for the 2°-1°-2° pattern of cell fates expressed by those Pn.p cells that have been induced by the anchor cell. The evidence for a graded inductive signal is suggestive but not compelling. The evidence for a lateral signal is based primarily on studies of *lin-15* mutants. The relative contribution of these two signals to wild-type vulval development is unknown.

We postulate that both the inductive and the lateral signalling systems function during normal vulval development. Specifically, we propose that the anchor cell has two roles in establishing the spatial pattern of Pn.p cell fates (Fig. 5): first, it induces three cells, making them competent to express the 1° and 2° fates; second, it biases the fates of these cells so that the cell closest to the anchor cell is most likely to express the 1° fate. In addition, induction by the anchor cell could activate the lateral signalling system (for example, the anchor cell could induce expression of either the lateral signal or of *lin-12*), because expression of the 2° fate requires inductive signalling and can be caused by lateral signalling. Once induced, the Pn.p cells interact through lateral signalling, which amplifies small differences in the strength of the inducing signal received by neighbouring cells, so helping to establish the precise spatial pattern of Pn.p cell fates.

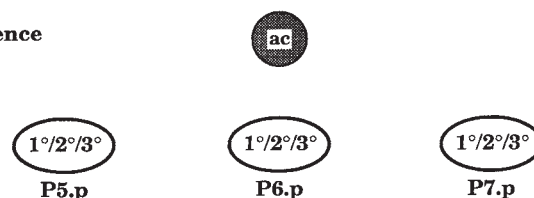
In the model shown in Fig. 4, the inductive signalling system involves the anchor cell, the *hyp7* syncytial hypoderm and the Pn.p cells. The *Vul* and *Muv* genes discussed above, including *let-23*, *let-60*, *lin-10* and *lin-15*, function in this system. On the other hand, *lin-12* acts in the lateral signalling system and conceivably could also act in a 2°-specific inductive signalling system.

Cell interactions control a transcription factor

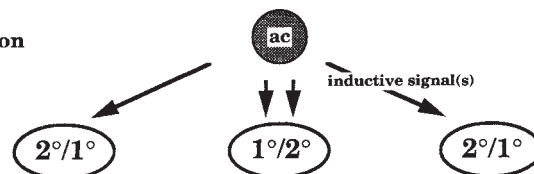
Once the fate of each Pn.p cell is determined by cell interactions to be 1°, 2° or 3°, that fate must be expressed. Three genes have been identified that are necessary for the expression of the 2° fate^{11,45}. One of these genes, *lin-11*, controls the asymmetry of the divisions of cells expressing the 2° fate. These cells normally produce one daughter that generates descendants of two distinct cell types, called T and N, and one daughter that generates descendants of another cell type, called L (see Fig. 2c)⁴. In mutants lacking *lin-11* function, 2° Pn.p cells produce descendants only of the L cell type^{11,19,46}. Thus, *lin-11* normally acts in vulval development to make the two daughters of cells expressing the 2° fate different from one another.

The *lin-11* gene seems to encode a protein with a homeo-domain, a highly conserved motif that mediates DNA binding by a variety of transcription factors⁴⁶. The protein also contains two copies of a possible metal-binding domain known as the LIM motif, and is a member of a family of proteins conserved from nematodes to mammals⁴⁶⁻⁴⁹. Thus, *lin-12*, which causes

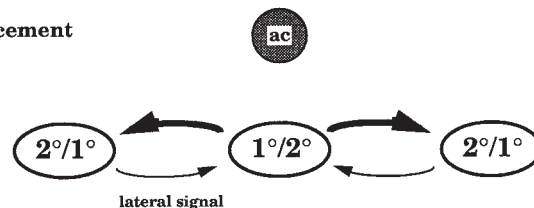
Competence



Induction



Reinforcement



Commitment

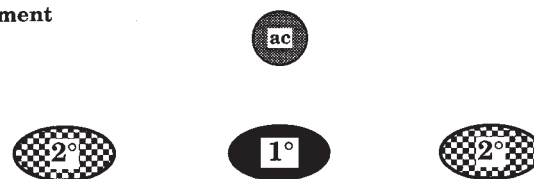


FIG. 5 Model for the series of events that occur while intercellular signals specify the fates of the Pn.p cells. Initially, P5.p, P6.p and P7.p are tripotent (competence). The anchor cell induces these three cells to express either the 1° or the 2° fate and possibly to activate the lateral signalling system (induction). The anchor cell also biases the nearest Pn.p cell toward the 1° fate, which causes this cell to generate more lateral signal, driving neighbouring Pn.p cells towards the 2° fate. Then either these neighbouring cells produce less lateral signal or the cell nearest the anchor cell becomes less responsive to the lateral signal, reinforcing the drive of the first cell towards the 1° fate (reinforcement). This process continues until the cell nearest the anchor cell becomes committed to the 1° fate and the neighbouring Pn.p cells become committed to the 2° fate (commitment).

the expression of the 2° fate, could activate a transcription factor that independently of subsequent cell interactions makes the daughters of the 2° Pn.p cells different from one another.

Inductive influence of vulval cells

During egg laying, the vulva is opened by a set of vulval muscles that are innervated by two classes of neurons, the serotonergic HSN neurons and the peptidergic VC neurons^{8,50,51}. Both the HSN and VC neurons have axonal branches that provide synaptic input to the vulval muscles. In *dig-1* animals with anteriorly displaced gonads, the VC neurons⁸ and the HSN neurons⁵ branch at the site of the displaced vulva. In Vul animals, neither the VC nor the HSN neurons branch, whereas in Muv animals both the VC and the HSN neurons branch at the ectopic vulval-like structures⁸ (G. Garriga and H.R.H., unpublished results). Thus, vulval cells seem to induce the branching of the VC and HSN neurons.

Prospects

Current studies of vulval development aim to elucidate the molecular mechanisms of action of about 30 genes involved in the cell interactions that specify vulval cell fates. Molecular biological studies of these genes and their products should help define a molecular genetic pathway for vulval development. Further genetic studies should identify new genes as well as establish when and where specific genes act (complementing molecular findings about when and where their products appear), pathways of gene action, and aspects of the relationship between gene structure and function. The combination of these molecular and genetic studies should lead to a precise and comprehensive picture of cell signalling in vulval development.

The evidence that *let-23* and *lin-12* encode transmembrane receptors, that *let-60* encodes a Ras protein, and that *lin-11* encodes a transcription factor helps our thinking about how these genes function and how genes that act upstream or down-

stream of each of them might function. It is striking that at least four of the five genes that are involved in vulval development and that have been molecularly characterized encode proteins highly conserved from nematodes to mammals. Continued study of these genes should reveal not only their mechanisms of action in *C. elegans* vulval development but also aspects of their structure and function relevant to many other organisms. In addition, genes such as *lin-10*, which encodes a novel protein, could define new classes of molecules that function in signalling processes in other biological systems.

More generally, it seems likely that an understanding of both the molecular and cellular mechanisms responsible for the cell interactions of *C. elegans* vulval development will prove relevant to cell signalling throughout the biological world. Equivalence groups (sets of multipotent cells with fates determined by cell interactions between members of the set) have also been identified in leeches⁵², insects^{53,54} and ascidians⁵⁵. Members of the highly conserved *lin-12*/Notch protein family control cell signalling within equivalence groups in both nematodes and insects^{39,40}, and are also found in vertebrates⁴¹. The problems of permissive versus instructive induction, of sequential inductions, of how diffusible signals and signals mediated by direct cell contact act both separately and together, and of how spatial patterns of distinct cell types are reliably established, are fundamental to developmental biology (for example, see refs 56–58). In addition, given the role of cell signalling in humans not only in normal development but also in various oncogenic processes^{59,60}, the study of simple systems like *C. elegans* vulval development should have implications for the understanding of human disease. □

H. Robert Horvitz is in the Department of Biology, Howard Hughes Medical Institute, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA; Paul W. Sternberg is in the Division of Biology, Howard Hughes Medical Institute, California Institute of Technology, Pasadena, California 91125, USA.

- Wood, W. B. et al. in *The Nematode Caenorhabditis elegans* (Cold Spring Harbor Laboratory, New York, 1988).
- Sulston, J. E. & Horvitz, H. R. *Dev. Biol.* **56**, 110–156 (1977).
- Sulston, J. E. & White, J. G. *Dev. Biol.* **78**, 577–597 (1980).
- Sternberg, P. W. & Horvitz, H. R. *Cell* **44**, 761–772 (1986).
- Thomas, J. Stern, M. & Horvitz, H. R. *Cell* **62**, 1041–1052 (1990).
- Sulston, J. E., Schierenberg, E., White, J. G. & Thomson, N. *Dev. Biol.* **100**, 64–119 (1983).
- White, J. G. in *The Nematode Caenorhabditis elegans* (eds Wood, W. B. et al.) (Cold Spring Harbor Laboratory, New York, 1988).
- Li, C. & Chalfie, M. *Neuron* **4**, 681–695 (1990).
- Kimble, J. *Dev. Biol.* **87**, 286–300 (1981).
- Ferguson, E. & Horvitz, H. R. *Genetics* **110**, 17–72 (1985).
- Ferguson, E., Sternberg, P. W. & Horvitz, H. R. *Nature* **326**, 259–267 (1987).
- Herman, R. K. & Hedgecock, E. M. *Nature* **348**, 169–171 (1990).
- Herman, R. K. *Genetics* **108**, 165–180 (1984).
- Sternberg, P. W. *Nature* **335**, 551–554 (1988).
- Han, M., Aroian, R. & Sternberg, P. W. *Genetics* **126**, 899–913 (1990).
- Horvitz, H. R. & Sulston, J. E. *Genetics* **96**, 435–454 (1980).
- Sulston, J. E. & Horvitz, H. R. *Dev. Biol.* **82**, 41–55 (1981).
- Greenwald, I. S., Sternberg, P. W. & Horvitz, H. R. *Cell* **34**, 435–444 (1983).
- Ferguson, E. & Horvitz, H. R. *Genetics* **123**, 109–121 (1989).
- Kim, S. & Horvitz, H. R. *Genes Dev.* **4**, 357–371 (1990).
- Beitel, G., Clark, S. & Horvitz, H. R. *Nature* **348**, 503–509 (1990).
- Han, M. & Sternberg, P. W. *Cell* **63**, 921–993 (1990).
- Barbacid, M. A. *Rev. Biochem.* **56**, 779–827 (1987).
- Toda, T. et al. *Cell* **40**, 27–36 (1985).
- Downward, J., Graves, J., Warne, P., Rayter, S. & Cantrell, D. *Nature* **346**, 719–723 (1990).
- Bourne, H. R., Sanders, D. A. & McCormick, F. *Nature* **348**, 125–132 (1990).
- Aroian, R., Koga, M., Mendel, J., Ohshima, Y. & Sternberg, P. W. *Nature* **348**, 693–699 (1990).
- Ullrich, A. & Schlessinger, J. *Cell* **61**, 203–212 (1990).
- Schubach, T. *Cell* **49**, 699–707 (1987).
- Price, J., Clifford, R. & Schubach, T. *Cell* **56**, 1085–1092 (1989).
- Schejter, E. & Shilo, B.-Z. *Cell* **56**, 1093–1104 (1989).
- Baker, N. & Rubin, G. *Nature* **340**, 150–153 (1989).
- Sternberg, P. W. & Horvitz, H. R. *Cell* **58**, 679–693 (1989).
- Seydoux, G. & Greenwald, I. S. *Cell* **57**, 1237–1245 (1989).
- Greenwald, I. S. *Cell* **43**, 583–590 (1985).
- Yochem, J., Weston, K. & Greenwald, I. S. *Nature* **335**, 547–550 (1988).
- Andrews, B. & Herskowitz, I. *J. Biol. Chem.* **265**, 14057–14060 (1990).
- Lux, S., John, K. & Bennett, W. *Nature* **344**, 36–42 (1990).
- Greenwald, I. S. *Trends Genet.* **5**, 237–241 (1989).
- Greenspan, R. *New Biol.* **2**, 595–600 (1990).
- Coffmann, C., Harris, W. & Kintner, C. *Science* **249**, 1438–1441 (1990).
- Kidd, S., Maylies, M., Gasic, G. & Young, M. *Genes Dev.* **3**, 1113–1129 (1989).
- Fehon, R. et al. *Cell* **61**, 523–534 (1990).
- Greenwald, I. S. & Seydoux, G. *Nature* **346**, 197–199 (1990).
- Sternberg, P. W. & Horvitz, H. R. *Dev. Biol.* **130**, 67–73 (1988).
- Frey, G., Kim, S. & Horvitz, H. R. *Nature* **344**, 876–879 (1990).
- Way, J. & Chalfie, M. *Cell* **54**, 5–16 (1988).
- Karlsson, O., Thor, S., Norberg, T., Ohlsson, H. & Edlund, T. *Nature* **344**, 879–882 (1990).
- Rabbitts, T. & Boehm, T. *Nature* **348**, 418 (1990).
- White, J. G., Southgate, E., Thomson, N. & Brenner, S. *Phil. Trans. R. Soc. B* **314**, 1–340 (1986).
- Desai, C., Garriga, G., McIntire, S. & Horvitz, H. R. *Nature* **336**, 638–646 (1988).
- Shankland, M. *Trends Genet.* **3**, 314–319 (1987).
- Doe, C. & Goodman, C. *Dev. Biol.* **111**, 206–219 (1985).
- Simpson, P. *Development* **109**, 509–519 (1990).
- Nishida, H. & Satoh, N. *Dev. Biol.* **132**, 355–367 (1989).
- Lewis, J., Slack, M. & Wolpert, L. *J. theor. Biol.* **65**, 575–590 (1977).
- Gurdon, J. *Development* **99**, 285–306 (1987).
- Banerjee, U. & Zipursky, L. *Neuron* **4**, 177–187 (1990).
- Weinberg, R. *J. NIH Res.* **3**, 45–48 (1991).
- Cantley, L. et al. *Cell* **64**, 281–302 (1991).

ACKNOWLEDGEMENTS. We thank the members of our laboratories, particularly C. Bargmann and E. Jorgensen, and L. Avery, M. Chalfie, E. Davidson, C. Ferguson, I. Greenwald, B. Herman, I. Herskowitz, C. Kenyon, S. Kim, J. Kimble, K. Struhl and J. White for many helpful suggestions. The work done in our laboratories was supported by the US Public Health Service, H.R.H. and P.W.S. are investigators of the Howard Hughes Medical Institute.

Genetic analysis of autoimmune type 1 diabetes mellitus in mice

John A. Todd, Timothy J. Aitman, Richard J. Cornall, Soumitra Ghosh, Jennifer R. S. Hall, Catherine M. Hearne, Andrew M. Knight*, Jennifer M. Love, Marcia A. McAleer, Jan-Bas Prins, Nanda Rodrigues, Mark Lathrop†, Alison Pressey‡, Nicole H. DeLarato‡, Laurence B. Peterson§ & Linda S. Wicker‡

Nuffield Department of Surgery, John Radcliffe Hospital, Headington, Oxford OX3 9DU, UK

* Transplantation Biology, Clinical Research Centre, Watford Road, Harrow, Middlesex HA1 3UJ, UK

† CEPH, 27 rue Juliette Dodu, Paris 75010, France

‡ Autoimmune Diseases Research and § Department of Cellular and Molecular Pharmacology, Merck Sharp & Dohme Research Laboratories, Rahway, New Jersey 07065, USA

Two genes, *Idd-3* and *Idd-4*, that influence the onset of autoimmune type 1 diabetes in the nonobese diabetic mouse have been located on chromosomes 3 and 11, outside the chromosome 17 major histocompatibility complex. A genetic map of the mouse genome, analysed using the polymerase chain reaction, has been assembled specifically for the study. On the basis of comparative maps of the mouse and human genomes, the homologue of *Idd-3* may reside on human chromosomes 1 or 4 and *Idd-4* on chromosome 17.

ONE route to understanding the aetiology and pathogenesis of disorders such as hypertension, schizophrenia and diabetes, which are caused both by genetic and by environmental factors, is to identify genes that predispose individuals to these diseases. Disease-associated genetic markers can be used to identify high-risk individuals and therapeutic strategies may be developed on the basis of the biochemical functions of the disease gene products. Human type 1 diabetes is an organ-specific autoimmune disease of the pancreas affecting 0.2–0.3% of populations of European descent that causes tissue damage, leading to blindness and kidney failure, and reduces life expectancy^{1–3}. Affected individuals require daily injections of insulin. Over a long and variable period, cells of the immune system infiltrate the pancreas, invade the islets of Langerhans—a process called insulinitis—and selectively destroy the insulin-producing β cells^{2,3}.

Identical twins of affected individuals have a 36% risk of developing type 1 diabetes⁴ compared with an average 6% risk for other siblings⁵. Certain alleles of at least three major histocompatibility complex (MHC) class II genes, *HLA-DQA1*, *-DQB1* and *-DRB1* on chromosome 6, correlate with predisposition to type 1 diabetes⁶. The insulin gene region on human chromosome 11p is also associated with disease susceptibility⁷. But there is no reliable estimate of the number of genes involved or how many different genotypic combinations give rise to the same clinical diabetic phenotype (referred to as genetic or locus heterogeneity)⁸.

The nonobese diabetic (NOD) mouse spontaneously develops insulin-dependent, type 1 diabetes with remarkable similarities to the human disorder⁹. Autoimmune islet-cell destruction is a shared characteristic^{2,3,9}, as are the appearance of autoantibodies to β -cell components^{2,3}, defects in T-cell activity^{10–12}, sensitivity of the disease to immunosuppression^{2,13} and the presence of susceptibility genes in the MHC^{6,14–17}. One difference is that NOD females have approximately twice the frequency of disease compared to males whereas in humans

males and females are equally affected. The environment also plays an important part in onset of diabetes in the NOD mouse⁶. The frequency of diabetes can be increased to 100% in germ-free conditions¹⁸. Conversely, viral infection of NOD mice reduces disease frequency^{19,20}.

A susceptibility gene, *Idd-1*, has been located in the murine MHC on chromosome 17 (refs 14–17, 21–23). *Idd-1* could be a gene complex with at least two susceptibility loci, that is, the class II genes *I-A β* and *I-E α* . Transgenic mouse experiments suggest that the *I-A β* ⁸⁷ molecule (formerly known as *I-A β ^{nod}*), which is homologous to the human type 1 diabetes-associated allele HLA-DQB1*0302, influences disease development^{21–23}.

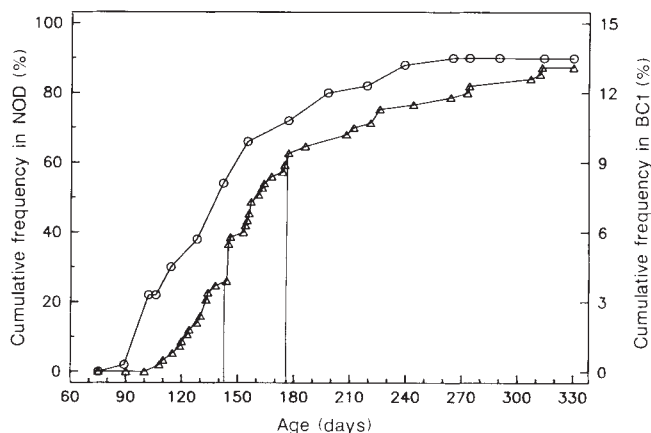


FIG. 1 The cumulative frequency distributions of diabetes in the parental NOD strain and combined data from the two backcrosses (B10.H-2^{b7} × NOD)_{F1} × NOD and NOD × (B10.H-2^{b7} × NOD)_{F1}. A cohort of female BC1 mice ($n=382$) were aged to 330 days and the cumulative frequency of diabetes compared to a cohort of NOD female mice ($n=50$) aged to 330 days. Most male progeny were sacrificed at weaning. To develop the B10.H-2^{b7} strain, B10 mice were outcrossed to NOD mice and resulting F_1 progeny were backcrossed to B10 as reported previously^{20,21}. Mice were intercrossed at the N6 generation to fix the NOD MHC on the B10 background. Three B10.H-2^{b7} females from the same litter were mated to NOD males and resulting F_1 mice were used to set up the backcrosses. All mice were bred and housed under specific pathogen-free conditions at Merck Sharp & Dohme, Rahway, New Jersey. Animals were monitored for elevations in urinary glucose using Tes-Tape (Eli Lilly, Indianapolis) and were classified as diabetic after producing consistent Tes-tape values of greater than 2⁺. Diagnosis of type 1 diabetes was confirmed by histology of the pancreas. Tissues were taken from mice classified as diabetic and also 97 non-diabetics (aged to at least 330 days) for histologic processing^{16,24–26}. Pancreata were fixed in neutral buffered formalin and processed for paraffin embedding. Tissue sections (4 μ m) were stained with haematoxylin and eosin. Histology of mice classified as diabetic with high levels of glucose in the urine showed massive destruction of islets. The vertical lines that intersect the BC1 cumulative frequency distribution result from division of the diabetics into thirds as described in Table 2. NOD, $\circ$; BC1, $\triangle$.

TABLE 1 A linkage map of the mouse genome and associations of markers with type 1 diabetes

Chromosome (location, cM)	Locus	Diabetics		Non- diabetics		$\chi^2 > 4$	Chromosome (location, cM)	Locus	Diabetics		Non- diabetics		$\chi^2 > 4$
		He	Ho	He	Ho				He	Ho	He	Ho	
1 (3)	<i>D1Nds4</i>	38	58	57	37	8.4	9 (24)	<i>Thy-1</i>	41	56	49	47	
1 (41)	<i>Bcl-2</i>	45	52	49	47		9 (29)	<i>Ncam</i>	39	58	48	49	
1 (42)	<i>D1Nds2</i>	44	49	50	47		9 (33)	<i>Cyp1a2</i>	39	58	49	48	
1 (48)	<i>D1Nds1</i>	50	47	49	46		9 (44)	<i>D9Nds2</i>	44	53	45	52	
1 (73)	<i>Crp</i>	57	40	45	51		9 (46)	<i>D9Nds1</i>	44	52	46	48	
2 (35)	<i>D2Nds1</i>	40	46	18	31	5.4	10 (29)	<i>D10Nds1</i>	47	50	49	46	
2 (46)	<i>B2m</i>	39	44	19	29		11 (10)	<i>Glns</i>	46	51	51	46	
3 (32)	<i>Il-2</i>	33	64	49	48		11 (42)	<i>Acrb</i>	31	66	51	46	
3 (53)	<i>D3Nds1</i>	17	80	54	43		11 (47)	<i>D11Nds1</i>	30	67	51	46	
3 (67)	<i>Tshb</i>	24	73	50	47		11 (52)	<i>Mpo</i>	36	61	53	44	
3 (86)	<i>Adh-1</i>	28	69	49	48	9.5	11 (68)	<i>Gfap</i>	36	61	51	46	4.7
4 (18)	<i>D4Nds3</i>	44	40	9	10		11 (71)	<i>Myla</i>	36	61	50	47	
4 (29)	<i>Mup-1</i>	43	43	21	28		12 (4)	<i>Odc</i>	54	43	47	50	
4 (30)	<i>Orm-1</i>	44	42	21	28		12 (45)	<i>Mtv-9</i>	36	41	13	16	
4 (62)	<i>D4Nds2</i>	40	56	55	40		13 (20)	<i>Hist1</i>	42	32	14	21	
4 (69)	<i>Lck</i>	44	53	52	41	4.6	13 (39)	<i>D13Nds1</i>	58	39	36	60	9.6
4 (95)	<i>Pnd</i>	11	20	25	15		13 (68)	<i>P198-13</i>	45	27	?	?	
5 (10)	<i>D5Nds1</i>	50	47	51	45		14 (8)	<i>Plau</i>	68	29	43	54	
5 (30)	<i>D5Nds2</i>	51	43	53	43		14 (27)	<i>Tcra</i>	61	36	42	52	
5 (46)	<i>Afp</i>	49	37	22	27		14 (38)	<i>Nfl</i>	52	45	45	47	
5 (94)	<i>Zp-3</i>	41	36	28	20	10.5	14 (42)	<i>Hpg</i>	55	42	47	49	4.1
6 (32)	<i>Ly-3</i>	38	48	22	27		15 (18)	<i>Myc</i>	38	59	48	46	
6 (68)	<i>Prp</i>	36	60	30	24		15 (24)	<i>D15Nds1</i>	37	60	50	45	
7 (6)	<i>Ckmm</i>	64	33	41	55		15 (27)	<i>Ly-6C</i>	38	59	51	45	
7 (27)	<i>Ngfg</i>	53	44	37	54		15 (49)	<i>Gdc-1</i>	32	48	23	19	
7 (48)	<i>D7Nds2</i>	42	53	38	58	4.6	15 (53)	<i>Hox-3</i>	40	57	52	44	
7 (64)	<i>Hbb</i>	39	55	44	50		16 (42)	<i>D16Nds2</i>	26	31	16	18	
8 (0)	<i>Polb</i>	43	43	21	28		18 (24)	<i>Fim-2</i>	37	40	21	17	
8 (35)	<i>Mt-2</i>	37	49	26	23		18 (29)	<i>li</i>	38	46	20	18	
							19 (35)	<i>Cyp2c</i>	43	37	17	21	
							X (23)	<i>Hprt</i>	29	28	25	14	
							X (39)	<i>DXNds3</i>	28	29	27	16	

Markers were ordered by minimizing double recombinants in the BC1 typing data with the aid of the program, GENELINK (X. Montagutelli, personal communication), by analysis of the strain distribution patterns of markers in recombinant inbred strain panels, using the program RIManager (K. Manly, personal communication) and by comparison with the published genetic map from the Jackson laboratory³⁸. Gene orders were confirmed by maximum likelihood analysis with the GMS program⁴⁶. The details of the linkage data will be published elsewhere. Map distances were calculated from recombination frequencies using Kosambi's formula, based on scoring of at least 70 animals from the backcrosses (both diabetics and non-diabetics). The most centromeric marker was used as an anchor locus if the microsatellite corresponded to a previously mapped gene. For chromosomes 1, 2, 3, 4 and 5, respectively, anchor markers were *Bcl-2*, *B2m*, *Tshb*, *Mup-1* and *Afp*. All markers except *B2m*, *Lck*, *Mtv-9*, *Hist1*, *P198*, *Ly-6*, *Fim-2* and *Cyp2c* (P450.2C)³⁷ are PCR-analysed microsatellites³³⁻³⁶. The sequences of PCR primers and methods have been described: *Bcl-2*, *Mup-1*, *Ly-3*, *Prp*, *Ngfg*, *Mt-2*, *Ncam*, *Cyp1a2*, *Glns*, *Acrb*, *Gfap*, *Myla*, *Plau*, *Tcra*, *Nfl*, *Myc*, *Gdc-1*, *Hox-3*, *Hprt* and *DXNds3* are dinucleotide repeat microsatellites³³; *Tshb*, *Afp*, *Thy-1*, *Odc* and *Zp-3* are mononucleotide repeats³⁴; *Crp*, *Adh-1*, *Orm-1*, *Pnd*, *Ckmm*, *Hbb*, *Polb*, *D9Nds1*, *Mpo*, *Hpg* and *li* are other microsatellites³⁵; *D1Nds1*, *D1Nds2*, *D2Nds1*, *D3Nds1*, *D4Nds2*, *D5Nds1*, *D5Nds2*, *D7Nds2*, *D9Nds2*, *D10Nds1*, *D11Nds1*, *D13Nds1*, *D15Nds1* and *D16Nds2* are dinucleotide repeat microsatellites randomly cloned from total NOD genomic DNA³⁶; *Il-2* is a further dinucleotide repeat microsatellite³⁵. *D1Nds4* is a dinucleotide repeat microsatellite with PCR primers, 5'-CTACATTTATCTACTCTCTAATC-3' and 5'-ATACTGTTGATACAGTCTGTAATG-3'. Allelic variations at *B2m* (ref. 47) and *Lck* (ref. 48) were detected by restriction enzyme digestion of PCR products and agarose gel electrophoresis: *B2m* PCR primers, 5'-CCGGAGAATGGGAAGCCGAACACTACTGAAC-3' and 5'-AGTTTTTAAGTCCACACAGATGGAGCGTCCA-3', product size 301 base pairs (bp), cut with *Bgl*I give fragments of 217 and 84 bp with B10 amplified DNA and a 301-bp product with NOD and *Lck* primers, 5'-GCAGATGGAATTCCTGTGCCA-3' and 5'-ACACACAGAGACATGAGATTGGAT-3', product size 340 bp, cut with *Hae*III give informative bands of 260 bp and 200 bp with B10- and NOD-amplified DNA, respectively. Three oligonucleotide primers were designed and used together in the PCR for *Ly-6C*: *Ly-6.1*, 5'-CAACCCATACCTCTATTAG-3', *Ly-6.2*, 5'-GAACTAACCACTACCCACAG-3' and *Ly-6.6*, 5'-TTGTTTTTGTCTTTT-TTGAGACA-3'. Standard RFLVs for *Mtv-9*, *Hist1*, *P198*, *Fim-2* and *Cyp2c* (P450.2C) were described previously³⁷ and we acknowledge G. Peters, W. F. Marzluff, E. DePlaen, S. Gisselbrecht and T. Friedberg, respectively, for probes. Because the congenic B10.H-2^{g7} females were used at the fifth backcross, it is expected that 3% of markers in the N6 breeders will be NOD type in addition to H-2. Two markers, *Pnd* and *D16Nds2* were thus affected in two and one, respectively, of the three breeders and this is the explanation for the reduction in numbers of animals scored for these two markers. He, heterozygous; Ho, NOD homozygous.

The relative contributions of the MHC and non-MHC genes to susceptibility to type 1 diabetes have been analysed by the use of experimental crosses and congenic mouse strains^{14-17,24-26}. A congenic strain NOD.B10.H-2^b (NOD.H-2^b), in which the H-2^{g7} of the NOD has been replaced with the H-2 region from the diabetes-resistant C57BL/10SnJ (B10) strain, does not develop insulinitis or diabetes demonstrating that the NOD MHC is essential for β -cell destruction (refs 24, 25). But neither destructive insulinitis nor diabetes are observed in a reciprocal

congenic strain, B10.NOD.H-2^{g7} (N6F2), referred to as B10.H-2^{g7}, showing that non-MHC genes are also required. These non-MHC genes appear to be phenotypically recessive because (B10.H-2^{g7} × NOD)F₁ hybrids are also resistant to insulinitis and diabetes^{24,25}.

In outcross-backcross studies of NOD with the diabetes-resistant strains C3H (ref. 14), NON (ref. 15), B10 (ref. 16) and SWR (ref. 17), the frequency of diabetes has been found to be about 5-15 times less than that for the parental NOD strain.

TABLE 2 Evidence for an association of chromosome 11 with diabetes

Locus		Diabetics								Non- diabetics
		94-143		Age-of-onset (days)				177-420		
		He:Ho	χ^2	He:Ho	144-175		He:Ho	χ^2		
1 (3)	<i>D1Nds4</i>	11:21	6.6	14:18				13:19		57:37
3 (32)	<i>Il-2</i>	6:26	9.9	9:23	4.9			18:15		49:48
3 (53)	<i>D3Nds1</i>	4:28	18.1	7:25	11.0			6:27	13.9	54:43
3 (67)	<i>Tshb</i>	9:23	5.3	9:23	5.3			6:27	11.2	50:47
3 (86)	<i>Adh-1</i>	9:23	4.9	11:21				8:25	6.9	49:48
6 (68)	<i>Prp</i>	15:17		11:21				10:22	4.8	30:24
7 (6)	<i>Ckmm</i>	23:9	8.2	23:9	8.2			18:15		41:55
7 (27)	<i>Ngfg</i>	20:12	4.5	17:15				16:17		37:54
9 (24)	<i>Thy-1</i>	9:23	5.1	13:19				19:14		49:47
9 (29)	<i>Ncam</i>	8:24	5.9	12:20				19:14		48:49
11 (42)	<i>Acrb</i>	6:26	11.2	11:21				14:19		51:46
11 (47)	<i>D11Nds1</i>	4:28	15.8	11:21				15:18		51:46
11 (52)	<i>Mpo</i>	5:27	14.8	13:19				18:15		53:44
11 (68)	<i>Gfap</i>	7:25	9.2	13:19				16:17		51:46
11 (71)	<i>Myla</i>	7:25	8.6	13:19				16:17		50:47
12 (4)	<i>Odc</i>	25:7	8.6	16:16				13:20		47:50
13 (39)	<i>D13Nds1</i>	17:15		22:10	9.5			19:14	4.0	36:60
14 (8)	<i>Plau</i>	20:12		26:6	13.2			22:11	4.9	43:54
14 (27)	<i>Tcra</i>	19:13		21:11	4.2			21:12		42:52
15 (49)	<i>Gdc-1</i>	14:16		8:21	5.1			10:11		23:19
15 (53)	<i>Hox-3</i>	16:16		9:23	6.5			15:18		52:44

Diabetic animals, arranged according to age-of-onset, were divided into thirds: 32 animals with age-of-onset between 94 and 143 days, 32 animals at 144-175 days and 33 animals at 177-420 days. This division of the diabetics is shown graphically in Fig. 1 by vertical lines. Genotype frequency differences in each group compared to the non-diabetics were evaluated by χ^2 tests in 2×2 contingency tables. Only $\chi^2 > 4$ values are shown. Note the significant associations of *D11Nds1* and *Mpo*. The markers and their locations are as described in Table 1. He, heterozygous; Ho, homozygous.

These data suggest that at least two phenotypically recessive genes unlinked to *Idd-1* are necessary for disease onset. One of these non-MHC genes may control the frequency and severity of insulinitis and a second might influence the progression from insulinitis to clinical onset of diabetes¹⁶. A non-MHC gene, designated *Idd-2* has been tentatively assigned to mouse chromosome 9 (ref. 15).

To map genes that influence multifactorial traits systematically, two main reagents are required: a linkage map covering most of the genome of the relevant species^{27,28} and large numbers of both affected and unaffected progeny, generated by the crossing of susceptible and nonsusceptible individuals²⁷⁻²⁹. Neither of these criteria have been comprehensively fulfilled in the study of diabetes. We have thus assembled a genetic map of the mouse genome with newly characterized DNA markers, called microsatellites³⁰⁻³⁶. By analysis of the associations of these markers with disease in a large backcross to NOD with the diabetes-resistant (B10.H-2^{g7} × NOD)F₁, we have mapped susceptibility genes for type 1 diabetes to chromosomes 3 and 11. Histopathological examination of islets from non-diabetic backcross mice indicates that the chromosome 3 gene contributes to the frequency and severity of insulinitis. We have also found that the age at which diabetes is diagnosed has a substantial effect on the ability to detect the chromosome 11 association.

Murine type 1 diabetes is caused by alleles of at least three unlinked genes, one of which has variable, age-dependent penetrance. The human homologues of these genes may influence susceptibility to human type 1 diabetes.

Backcrosses

An important design feature of our study was the elimination of *Idd-1* as a variable in the backcross progeny by using the congenic B10.H-2^{g7} strain. Two reciprocal first backcross generations (BC1) were produced. (B10.H-2^{g7} × NOD)F₁ × NOD, generated 59/430 (13.7%) female and 4/142 (2.8%) male diabetic mice and NOD × (B10.H-2^{g7} × NOD)F₁ generated 42/282 (14.9%) female and 2/66 (3.0%) male diabetic mice. There was no significant difference in the frequency of type 1 diabetes between the two backcrosses indicating that the NOD X chromo-

somes does not carry a major, recessive susceptibility gene. Ninety seven out of 101 female diabetic BC1 progeny were analysed. Data from male diabetics are not included in this study. Figure 1 shows the cumulative frequency distribution of diabetes up to 330 days of the female parental NOD strain and of the female first backcross generations. The cumulative frequencies of diabetes at 330 days were 90% and 13% in the NOD parental and BC1 female mice, respectively. The frequency of diabetes in males in the BC1 generations was low at 3% (6/208) compared with the parental male NOD strain at 50% (data not shown).

Microsatellite genetic map

We developed 53 variant, locus-specific microsatellites³³⁻³⁶ that give, in combination with eight restriction fragment length variants (RFLV)³⁷, at least one marker per chromosome. Microsatellite markers on chromosome 17 (ref. 33) were not informative in the backcrosses because the MHC region was fixed. The Y chromosome could not be followed because only females were analysed and all male BC1 mice carry the NOD Y chromosome. Microsatellites contain tandem repeats of simple sequences such as the dinucleotides (CA)_n and (GA)_n ($n > 10$). These are highly variable between inbred strains of mice, and size differences between alleles are conveniently determined using the polymerase chain reaction (PCR) and gel electrophoresis³³⁻³⁶. The chromosomal locations and names of the markers are listed in Table 1. The map distances and order of the markers are consistent with the published genetic map of the mouse genome³⁸.

Chromosome 3

The associations of alleles of markers and type 1 diabetes were evaluated by a χ^2 test of gene frequencies in diabetic and non-diabetic animals (Table 1). By this test, only two markers have significant associations with diabetes: *D3Nds1* ($\chi^2 = 30.4$) and *Tshb* ($\chi^2 = 14.8$) on chromosome 3. The segregation of the markers in the non-diabetics did not differ significantly from random expectations. Therefore the associations are diabetes-specific. Associations were judged to be significant when the χ^2

exceeded 13.8. This critical value was chosen because it is numerically the same value as 3 on the logarithm of the odds (lod) scale³⁹, and exceeds the value necessary to assure that the proportion of true linkages among significant linkage is greater than 95% when testing randomly chosen genetic loci in the mouse genome (Neumann, personal communication). Of 97 diabetics, 80 (82%) are NOD homozygous for *D3Nds1* compared with 43/97 (43%) for non-diabetics (Table 1), implying that the NOD allele of the chromosome 3 disease-associated gene (*Idd-3*) is a positive susceptibility determinant. The chromosome 3 χ^2 values in Table 1 indicate that a susceptibility gene (or genes) probably lies between *Il-2* ($\chi^2 = 5.4$; about 32 centimorgans (cM) from the centromere) and *Tshb* ($\chi^2 = 14.8$; about 67 cM from the centromere). Nine out of 97 diabetics are heterozygous at *Il-2*, *D3Nds1* and *Tshb*. Because the possibility of all nine of these mice having double recombination events between *Il-2* and *D3Nds1* or *Tshb* and *D3Nds1* is very small, it is unlikely that any of the nine heterozygotes are actually NOD homozygous at the disease locus. This indicates that diabetes in this cross cannot be viewed as a simple recessive trait.

Chromosome 11

To evaluate the possibility of an influence of age on genetic susceptibility, the diabetic animals were ordered according to ascending age-of-onset and then divided into thirds so that there were roughly equal numbers in each group (Fig. 1 and Table 2). This division was not based on any particular choice of age-of-onset or genotype. Two adjacent loci on chromosome 11, *D11Nds1* ($\chi^2 = 15.8$) and *Mpo* ($\chi^2 = 14.8$), were significantly associated with diabetes only when the group of diabetics with the youngest age-of-onset (94–143 days) was compared with the non-diabetics (Table 2). The NOD allele of this gene, designated *Idd-4*, is associated with diabetes. As was found for *Idd-3* (this study) and *Idd-1*^{16,17,26}, at least 4/32 early-onset diabetics were likely to be heterozygous at *Idd-4* (data not shown and Fig. 3), indicating that the penetrance of the NOD allele of *Idd-4* is influenced by other genetic and environmental factors.

Location analysis

Estimates for the locations of *Idd-3* and *Idd-4* on chromosomes 3 and 11 were determined by multilocus linkage analysis, incorporating possible age-of-onset effects (see legend to Fig. 2). The one-, two- and three-lod-unit support intervals shown in Fig. 2 represent regions within which there is a different likelihood of

TABLE 3 *Idd-3* influences the frequency and severity of insulinitis

Genotype	Insulinitis		
	None	Mild	Moderate or severe
<i>D3Nds1</i> Ho (n=43)	13 (30%)	5 (12%)	25 (58%)
<i>D3Nds1</i> He (n=54)	37 (68%)	4 (7%)	13 (24%)
$\chi^2 > 4$	14.0		11.7
P value	<0.001		<0.001

Histological analysis was done on pancreata from 210–330-day-old non-diabetic backcross mice described in the legend to Fig. 1. Histology was scored in the following categories^{16,24–26}: 0, no evidence of lymphocytes in pancreas; 1, some periductal lymphocyte infiltration; 2, periislet infiltration, no insulinitis; 3, very mild insulinitis in some islets with no reduction in islet mass; 3b, extensive insulinitis with significant islet mass remaining; 3b*, extensive insulinitis with significant reduction in islet mass; 4*, as 3b*, but only residual islets remaining. For comparative analysis, these categories were combined as follows: 0, 1, 2, no insulinitis; 3, mild insulinitis; 3b, moderate insulinitis; 3b* + 4*, severe insulinitis. Differences in the frequencies of animals which were NOD homozygous (Ho) and heterozygous (He) at *D3Nds1* in the insulinitis categories were evaluated by χ^2 tests in 2 × 2 contingency tables.

the gene being located, as determined by the maximum likelihood estimates.

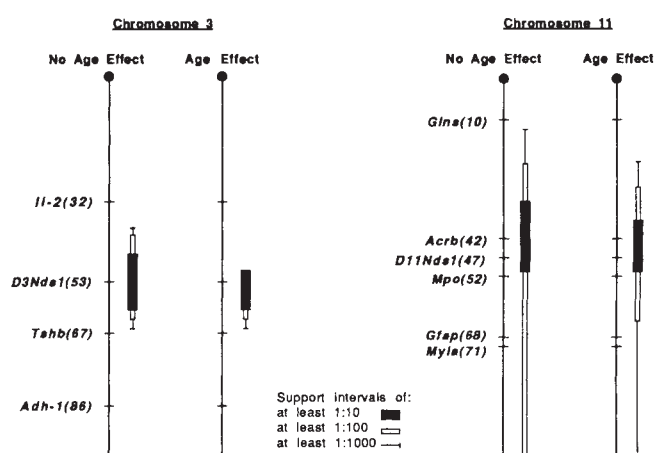
Idd-3 is localized to an interval on chromosome 3, flanked by *Il-2* and *Tshb* and containing *D3Nds1*. Alternatives with *Idd-3* placed outside this interval were rejected if the relative likelihood ratios were greater than 1,000:1. Although the maximum likelihood estimate for the location of *Idd-4* is between *Acrb* and *Mpo*, near *D11Nds1*, alternative placements in other intervals on chromosome 11 are possible with relative likelihood ratios between 3:1 and 1,000:1 (data not shown and Fig. 2). The inclusion of data on the age-of-onset of diabetes narrows the support intervals for both *Idd-3* and *Idd-4* (Fig. 2).

Genetic control of insulinitis

As we have shown that *Idd-3* and *Idd-4* are associated with type 1 diabetes their effects may be evident in the development of insulinitis, which is an essential stage in the onset of overt disease^{2,9}. Histological analysis^{16,24–26} of the pancreata from all 97 non-diabetic mice at an age of 210–330 days showed that 38/97 (39%) had moderate to severe insulinitis, 9/97 (9%) had mild insulinitis and 50/97 (51%) had no insulinitis (Table 3). Significantly more insulinitis-free animals were heterozygous than

FIG. 2 Maximum likelihood position estimates and support intervals for the locations of *Idd-3* and *Idd-4* on chromosomes 3 and 11, respectively. The location of the susceptibility gene is shown as the chromosome region within which there is at least a 1 in 10, 1 in 100 and 1 in 1,000 likelihood of finding the gene by maximum likelihood estimate.

METHODS. Support was evaluated under a genetic model which assumed a single susceptibility gene with unknown penetrance on the chromosome containing the markers linked to diabetes. Only diabetic animals were included in the location analysis. Conditional probabilities for a diabetic animal to be homozygous (NN) or heterozygous (NB) were estimated for each location or interval examined. The age of animals at diabetes onset was incorporated as a phenotype under the assumption of a logistic⁴⁹ age-of-onset function of the form, $P(NN|d) = \exp(\alpha + \beta * a) / (1 + \exp(\alpha + \beta * a))$ where $P(NN|d)$ is the conditional probability of NN given diabetes (d), a is the age at onset of diabetes, and α and β are unknown parameters. The hypothesis $\alpha = \beta = 0$ corresponds to $P(NN|d) = P(NB|d) = 0.5$, or to the absence of linkage, and $\beta = 0$ corresponds to the absence of an age effect. Likelihood ratio statistics were calculated for the evaluation of different placements of the susceptibility loci. Two approaches were used. First, we used a modification of the location score method⁵⁰, in which the recombination frequencies between the markers were fixed at the values calculated during the construction of the map, and the placement of the susceptibility locus was varied throughout the chromosome. New estimates for α and β were obtained for each map location that was examined. Second, we estimated the recombination rates, and the parameters α and β with the susceptibility gene placed in each interval of the map. The two approaches led to identical conclusions regarding the relative likelihoods



for placements of the susceptibility loci in different intervals of the map. Support intervals are shown as calculated from the modified location score method. Each marker was scored on all the available diabetic ($n=97$) and non-diabetic ($n=97$) female BC1 progeny. The numbers in parentheses are the distances in cM from the centromere (Table 1), which is shown as a circle on the top of the chromosome stick diagram.

homozygous at *D3Nds1* (compared to animals with insulinitis $\chi^2 = 14.0$; $P < 0.001$). Animals with moderate to severe insulinitis were more often homozygous at *D3Nds1* than heterozygous compared to animals with mild or no insulinitis ($\chi^2 = 11.7$; $P < 0.001$; Table 3). These data demonstrate that the NOD allele of *Idd-3* may contribute to the development and severity of insulinitis. No other marker or genotype, including those on chromosome 11, correlated significantly with insulinitis in these animals ($P > 0.001$; data not shown).

Discussion

Idd-3, which is near the chromosome 3 marker *D3Nds1* (Fig. 2), affects both insulinitis (Table 3) and diabetes (Table 1). This may be the susceptibility gene (or gene complex) that was proposed previously to be essential for the development of insulinitis¹⁶. Evidence from an F_2 cross of $(B10.H-2^{87} \times NOD)F_1 \times (B10.H-2^{87} \times NOD)F_1$ also indicates that *Idd-3* is an important factor for insulinitis. Only 1/24 F_2 progeny with insulinitis was B10 homozygous at *D3Nds1* and this animal was heterozygous at *Il-2* suggesting that it may also be heterozygous at *Idd-3* (data not shown). *Idd-1* influences diabetes susceptibility in all crosses with NOD so far reported¹⁴⁻¹⁷. To test the possibility that *Idd-3* is generally important in susceptibility to diabetes, a second BC1 to NOD with the diabetes-resistant strain, B6.PL-*Thy-1*^a/Cy was analysed. *Idd-3* also appears to influence diabetes onset in this cross because only 1/13 diabetic $(NOD \times B6.PL)F_1 \times NOD$ progeny was heterozygous at *Il-2*, *D3Nds1* and *Tshb* (unpublished data).

A second non-MHC gene, *Idd-4* maps close to *D11Nds1* (Tables 1 and 2, Fig. 2). Although *Idd-4* shows no association with insulinitis in the non-diabetic progeny (data not shown), *Idd-4* may influence the frequency of insulinitis. This association may be apparent in animals scored for insulinitis at an age less than 144 days because the association of *Idd-4* with diabetes depends on the age of the animal (Table 2 and Fig. 2). Another function for *Idd-4* in this genetic background may be to control the progression of severe insulinitis to overt diabetes¹⁶. If similar genetic heterogeneity related to age-of-onset is found in human type 1 diabetes, then this will have implications for the design of linkage studies of the human disease. For an immunological disease, age-dependent effects might be predicted because the immune system changes over time. For example, in mice deletion of T cells bearing $V\beta 14$ and $V\beta 11$ antigen receptors takes up to 150 days⁴⁰, which corresponds to the age-of-onset of diabetes at which we can no longer detect the association of *Idd-4*.

Diabetic BC1 progeny that were heterozygous at *Idd-1* have been observed^{17,26}. Consequently, it was proposed that the NOD

allele of *Idd-1* is dominant with incomplete penetrance^{16,17,24-26}. Similarly, we have found diabetic BC1 progeny which are likely to be heterozygotes at *Idd-3* (rows 2-5 of Fig. 3) or *Idd-4* (rows 6-8 of Fig. 3) but none is heterozygous at both loci (first row of Fig. 3). The proposed interaction between *Idd-3* and *Idd-4* is not statistically significant but the analysis raises the possibility that the influence of a predisposing allele of one gene may depend on alleles of one or more unlinked genes. Furthermore, in a case of locus heterogeneity where two (or more) susceptibility genes have the same effect, it may be possible to detect significant associations of either gene with disease only when the genotypes at both loci are analysed simultaneously. An example of locus heterogeneity has been described for the human disease, tuberous sclerosis⁴¹.

Investigation of the relative contributions of *Idd-3* and *Idd-4* to the pathogenesis of diabetes and their role in autoimmunity awaits development of NOD and B10 congenic strains. Congenic strains will allow analysis of the specific role of the *Idd-3* and *Idd-4* genes in insulinitis and diabetes and fine mapping of the diabetogenic mutations.

A locus, designated *Idd-2*, near *Thy-1* on chromosome 9 may influence diabetes in a $(NOD \times NON)F_1 \times NOD$ backcross¹⁵. Statistical analysis of these previous data¹⁵ shows that the association of *Thy-1* with diabetes in 19 diabetic BC1 progeny has a $\chi^2 = 4.4$. Similarly, we found no significant association of any marker on chromosome 9 with diabetes (Tables 1 and 2) or insulinitis (data not shown). If we assume that the genotype frequencies for *Ncam* (Table 1) represent a true association with diabetes, about 600 diabetic and 1,000 non-diabetic BC1 progeny would have to be scored to obtain significant evidence for an association ($\chi^2 > 13.8$). Alternatively, analysis of an F_2 cross may be a more efficient method of detecting dominant genes with variable penetrance. The effects of *Idd-1*, -2, -3 and -4 may differ depending on the diabetes-resistant strain used.

Because diabetes is not inherited in a simple recessive manner in this backcross, standard calculations to estimate the proportion of the genome covered by the markers in Table 1 are inappropriate⁴². Empirically, we can estimate from Table 1 that a marker should be within 20 cM of a locus which influences susceptibility with equivalent effects to those of *Idd-3* and *Idd-4*, to give a $\chi^2 > 4$. A $\chi^2 > 4$ can be considered as a starting point for further experimentation, such as the generation of more markers, affected progeny and/or congenic strains. Using this 20 cM guideline, about 70% of the genome is covered here in terms of detecting genes with effects equivalent to *Idd-4*. If the frequency of diabetes in mice homozygous for the NOD alleles of *Idd-1*, -3 and -4 is 90%, as in NOD females, then the expected frequency of diabetes in our female BC1 progeny is 22.5%. We observed 13%, which suggests that at least one other gene is also involved. These genes may lie in the 30% of the genome that remains to be analysed and/or they may not be detectable using the numbers of BC1 progeny analysed in this study. For example, the association of the chromosome 1 marker, *D1Nds4* merits further analysis.

The T-cell activation gene, *Ly-6C*, is a candidate susceptibility locus for autoimmunity because it is mutated in NOD and NZB, which is a model for the autoimmune disease systemic lupus erythematosus^{43,44}. *Ly-6C* was not significantly associated with diabetes (Tables 1 and 2). Although the chromosome 14, 13 and 7 associations did not reach statistical significance (because of the stringent critical value for association that has been adopted in our study; Table 1) the results suggest that genes on these chromosomes could be involved in diabetes susceptibility. These associations were unexpected because homozygosity for NOD alleles is associated with resistance to diabetes relative to heterozygosity (Table 1) and imply that the B10 parent may contribute not only diabetes resistance but also susceptibility alleles.

Comparative mapping of the mouse and human genomes has revealed extensive regions of homology^{38,45}. For example, for

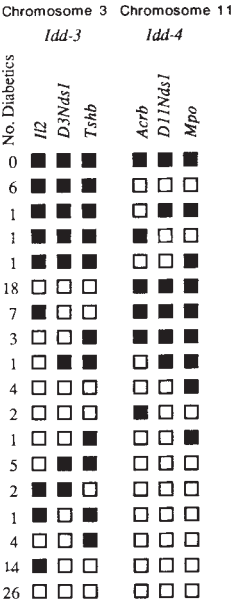


FIG. 3 Genotypes at the most likely locations for *Idd-3* and *Idd-4*. Each row is a particular genotype. Heterozygous, ■; NOD homozygous, □.

every gene that has been mapped to human chromosome 17, the homologue has been located on mouse chromosome 11. Predictions of the locations of disease loci in one species given their locations in the other species is an increasingly powerful

application of the mouse-human comparative maps. The homologues of *Idd-3* and *Idd-4* may therefore reside on human chromosomes 1 or 4, and 17, respectively, and are candidate susceptibility genes for human type 1 diabetes. □

Received 26 March; accepted 9 May 1991.

1. Metcalfe, M. A. & Baum, J. D. *Br. Med. J.* **302**, 443–447 (1991).
2. Castano, L. & Eisenbarth, G. S. *A. Rev. Immun.* **8**, 647–679 (1990).
3. Bottazzo, G. F. *et al. Br. Med. Bull.* **45**, 37–57 (1989).
4. Olmos, P. *et al. Diabetologia* **31**, 747–750 (1988).
5. Thomson, G. *et al. Am. J. hum. Genet.* **43**, 799–816 (1988).
6. Todd, J. A. *Immun. Today* **11**, 122–129 (1990).
7. Thomson, G., Robinson, W. P., Kuhnner, M. K., Joe, S. & Klitz, W. *Genet. Epidemiol.* **6**, 155–160 (1989).
8. Risch, N. *Am. J. hum. Genet.* **40**, 1–14 (1987).
9. Makino, S. *et al. Expl. Anim.* **29**, 1–13 (1980).
10. Serreze, D. V. & Leiter, E. H. *J. Immun.* **140**, 3801–3807 (1988).
11. Hatamori, N., Yokono, K., Hagata, M., Shii, K. & Baba, S. *Diabetes* **39**, 1070–1078 (1990).
12. Faustman, D., Eisenbarth, G., Daley, J. & Breitmeyer, J. *Diabetes* **38**, 1462–1468 (1989).
13. Sibley, R. K., Sutherland, D. E. R., Goetz, F. & Michael, A. F. *Lab. Invest.* **53**, 132–144 (1985).
14. Hattori, M. *et al. Science* **231**, 733–735 (1986).
15. Prochazka, M., Leiter, E. H., Serreze, D. V. & Coleman, D. L. *Science* **237**, 286–289 (1987).
16. Wicker, L. S. *et al. J. exp. Med.* **165**, 1639–1654 (1987).
17. Livingstone, A., Edwards, C. T., Shizura, J. A. & Fathman, C. G. *J. Immun.* **146**, 529–534 (1991).
18. Suzuki, T. *et al. in Immune-deficient Animals in Biomedical Research* (eds Rygaard, J., Groen, N. & Spang-Thomsen, M.), 112–116 (Karger, Basel, 1987).
19. Oldstone, M. B. A. *Science* **239**, 500–502 (1988).
20. Wilberz, S., Partke, H. J., Dagnaes-Hansen, F. & Herberg, L. *Diabetologia* **34**, 2–5 (1991).
21. Miyazaki, T. *et al. Nature* **345**, 722–724 (1990).
22. Slattery, R. M. *et al. Nature* **345**, 724–726 (1990).
23. Lund, T. *et al. Nature* **345**, 727–729 (1990).
24. Wicker, L. S., DeLarato, N. H., Pressey, A. & Peterson, L. B. in *Proceedings of the P & S Biomedical Sciences Symposium: Molecular Mechanisms of Immunological Self-Recognition* (eds Vogel, H. J. & Alt, S. W.) (Academic, New York, in the press).
25. Wicker, L. S., DeLarato, N. H., Pressey, A. & Peterson, L. B. in *Frontiers in Diabetes Research II: Lessons from Animal Diabetes III* (ed. Shafir, E.) (Smith-Gordon, in the press).
26. Wicker, L. S., Miller, B. J., Fischer, P. A., Pressey, A. & Peterson, L. B. *J. Immun.* **142**, 781–784 (1989).
27. White, R. *et al. Nature* **313**, 101–105 (1985).

28. Paterson, A. H. *et al. Genetics* **127**, 181–197 (1991).
29. Neumann, P. E. & Collins, R. L. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
30. Weber, J. L. & May, P. E. *Am. J. hum. Genet.* **44**, 388–396 (1989).
31. Litt, M. & Luty, J. A. *Am. J. hum. Genet.* **44**, 397–401 (1989).
32. Tautz, D. *Nucleic Acids Res.* **17**, 6463–6471 (1989).
33. Love, J. M., Knight, A. M., McAleer, M. A. & Todd, J. A. *Nucleic Acids Res.* **18**, 4123–4130 (1990).
34. Aitman, T. J., Hearne, C. M., McAleer, M. A. & Todd, J. A. *Mammalian Genome* (in the press).
35. Hearne, C. M. *et al. Mammalian Genome* (in the press).
36. Cornall, R. J., Aitman, T. J., Hearne, C. M. & Todd, J. A. *Genomics* (in the press).
37. Knight, A. M. & Dyson, P. J. *Molec. Cell. Probes* **4**, 497–504 (1990).
38. Hillyard, A. L., Doolittle, D. P., Davison, M. T. & Roderick, T. H. *Locus Map of the Mouse with Comparative Points of Human on Mouse* (The Jackson Laboratory, Bar Harbor, Maine, 1990).
39. Ott, J. in *Analysis of Human Genetic Linkage* (Johns Hopkins University Press, Baltimore, London, 1985).
40. Acha-Orbea, H. *et al. Nature* **350**, 207–211 (1991).
41. Janssen, L. A. J. *et al. Genomics* **8**, 237–242 (1990).
42. Carter, T. C. & Falconer, D. S. *J. Genet.* **50**, 307–323 (1951).
43. Philbrick, W. M., Maher, S. E., Bridgett, M. M. & Bothwell, A. L. M. *EMBO J.* **9**, 2485–2492 (1990).
44. Herold, K. C., Montag, A. G., Meyer, S. M., Wojcikowski, C. & Fitch, F. W. *Diabetes* **39**, 815–820 (1990).
45. Nadeau, J. H. *Trends Genet.* **5**, 82–86 (1989).
46. Lathrop, G. M. *et al. Genomics* **2**, 157–164 (1988).
47. Parnes, J. R. & Seidman, J. G. *Cell* **29**, 661–669 (1982).
48. Marth, J. D., Peet, R., Krebs, E. G. & Perlmutter, R. M. *Cell* **43**, 393–404 (1985).
49. Rice, J. & Risch, N. *Genet. Epidemiol.* **6**, 161–177 (1989).
50. Lathrop, G. M., Lalouel, J., Julier, C. & Ott, J. *Am. J. hum. Genet.* **37**, 482–498 (1985).

ACKNOWLEDGEMENTS. We thank our colleagues for critical review including J. Edwards, P. McShane and P. Neumann, who shared his ideas and methods. This work was supported by the Wellcome Trust, the Medical Research Council (as part of the UK Human Genome Mapping Project) the Juvenile Diabetes Foundation, the British Diabetes Association, the Eli Lilly Company and the Nuffield Foundation.

LETTERS TO NATURE

GT2318+620, a variable galactic source with a radio jet

A. R. Taylor*, P. C. Gregory†, N. Duric‡ & T. Tsutsumi‡

* Department of Physics and Astronomy, University of Calgary, 2500 University Drive NW, Calgary, Alberta, Canada T2N 1N4

† Department of Physics, University of British Columbia, Vancouver, British Columbia, Canada V6T 1W5

‡ Department of Physics and Astronomy, University of New Mexico, Albuquerque, New Mexico 87131, USA

A RECENT galactic survey with the National Radio Astronomy Observatory (NRAO) 91-m telescope revealed a small number of variable radio sources at low galactic latitude¹. Similar sources within the Galaxy, where they have been identified, are often associated with energetic objects such as X-ray binaries, pulsars, cataclysmic variables and flare stars. One of the new variable sources, GT2318+620, is within the error box for the Uhuru X-ray source 4U2316+61 (ref. 2), and our measurement of the neutral hydrogen absorption towards GT2318+620, reported here, indicates that it is a galactic object, at a distance of 3–6 kpc. We also present a high-resolution radio image of the source, which reveals an unresolved core with a jet-like feature extending on either side. GT2318+620 is coincident with a star of ~20th magnitude, and we suggest here that it is radio-emitting low-mass X-ray binary. The radio luminosity of GT2318+620 is markedly higher than that of Sco X-1, the only other low-mass X-ray binary to show a radio jet; in its radio brightness it resembles SS433, which is remarkable for its relativistic jet motion on arcsecond scales, thought to result from accretion onto a collapsed stellar object³.

GT2318+620 was classed as a possible variable source from three observations in August 1981 from a galactic plane survey¹. The source had a mean flux density of 23 ± 6 mJy at $\lambda = 6$ cm, and over a period of 12 days seemed to increase in flux density from 10 mJy to 40 mJy (ref. 1). At that time it was noted that the source lay within the error box of the X-ray source 4U2316+

61. To determine whether the radio source is of galactic or extragalactic origin, we obtained an absorption spectrum of the $\lambda = 21$ cm line of neutral hydrogen (H I) in the direction of the source with the Westerbork Synthesis radiotelescope (WSRT). The velocity structure of continuum absorption by interstellar hydrogen allows limits to be placed on the distance to the object. Radio images were obtained for a synthesized field containing GT2318+620 over 64 spectral line channels, each 4.12 km s^{-1} wide, covering a velocity range of $\sim 250 \text{ km s}^{-1}$. At the angular resolution of the instrument ($\sim 15''$) GT2318+620 appeared as a point-like radio source. Southeast of the variable source by 2.9 arcminutes is a strong double source. For comparison, we also measured an H I absorption spectrum in the direction of that source. The spectra are shown in Fig. 1. The absorption spectra are shown in the bottom panel, where we have plotted the optical depth of absorbing gas as a function of velocity in the frame of the local standard of rest. In the upper panel we plot the spectrum of H I emission in the same direction as GT2318+620 from the low-resolution survey of Weaver and Williams⁴, to show the velocities of the main interstellar features. A kinematic distance scale based on a flat galactic rotation curve outside the solar circle, with velocity $\Theta_{\odot} = 250 \text{ km s}^{-1}$, is also shown (the 'solar circle' is the circle around the centre of the Galaxy with radius equal to the distance of the Sun from the Galactic Centre).

The spectrum of the background source shows absorption features at 0, -10 and -50 km s^{-1} . The low-velocity gas from 0 to -10 km s^{-1} results largely from H I in the local Cygnus arm of the Galaxy. The H I at -50 km s^{-1} lies in the Perseus arm, which in this direction is ~ 6 kpc from the Sun⁵. Radiation from GT2318+620 is absorbed at 0 and -10 km s^{-1} but the feature at -50 km s^{-1} is absent, indicating that the radio source is located in front of the Perseus arm but beyond most of the gas in the local arm. This implies that the distance to the source is between ~ 3 and ~ 6 kpc.

On 4 November 1988, high-spatial-resolution observations of GT2318+620 were made at $\lambda = 20$ cm and $\lambda = 6$ cm with the 'A' configuration of the NRAO Very Large Array (VLA). The

observations were carried out with a bandwidth of 100 MHz, yielding a r.m.s. map noise of 0.1 and 0.07 mJy per beam at 20 and 6 cm respectively. The synthesized beams have full width at half maximum of $1.5''$ at 20 cm and $0.5''$ at 6 cm. The maps at each wavelength are shown in Fig. 2a and b. The 20-cm image reveals an unresolved core source with a jet-like feature extending to the north and south by $\sim 10''$. At 6 cm the emission is dominated by the core source, with a suggestion of a weak, arcsecond-scale jet straddling the core and orientated in the same direction as the 20-cm feature. The total flux density at 20 cm is 37 ± 2 mJy, which is significantly larger than the value of 15 ± 2 mJy measured at $\lambda = 21$ cm with the WSRT more than a year earlier. This difference may be due to variability or to the presence of additional extended emission which is not detectable by the 'A' configuration of the VLA. The region corresponding to the unresolved core at 20 cm has a flux density of 6.3 ± 0.3 mJy at 20 cm and 7.7 ± 0.3 mJy at 6 cm. The spectral index is thus flat or positive, indicating a compact emission region which is optically thick in parts. This is a characteristic common to known sources of highly variable radio emission. The lack of emission from the jet at 6 cm suggests an optically thin, nonthermal emission mechanism. Because of the high angular resolution, the 6-cm observations are not sensitive to low surface brightness, and therefore extended emission and additional flux may be present. We note, however, that GT2318+620 was not detected in the 6-cm survey of Condon *et al.*⁶, implying an upper limit to the 6-cm flux at that time of ~ 25 mJy. Further observations are required to determine the nature of the jet.

The coordinates of the core radio source are right ascension $\alpha(1950) = 23^{\text{h}} 18^{\text{m}} 22.57 \pm 0.02^{\text{s}}$, declination $\delta(1950) = 62^{\circ} 01' 07.3 \pm 0.15''$. We have searched the Palomar Sky Survey (PSS) plates for an optical counterpart to the radio source. Accurate positions of optical candidates were measured by solving for the transformation between plate coordinates (x, y)

and celestial coordinates (α, δ) using a set of nearby reference stars from the Smithsonian Astrophysical Observatory Catalog. An object of 20th magnitude is visible on the red PSS plate at coordinates $\alpha(1950) = 23^{\text{h}} 18^{\text{m}} 22.29 \pm 0.11^{\text{s}}$, $\delta(1950) = 62^{\circ} 01' 07.0 \pm 0.8''$. The difference in coordinates (radio – optical) are $\Delta\alpha = 1.9 \pm 0.8''$ and $\Delta\delta = 0.3 \pm 0.8''$. The quoted errors are 1σ . The radio and optical positions lie well within the 3σ error on the difference. The optical object is not visible on the blue PSS plate, suggesting either a high-reddening or an intrinsically red object.

Assuming a mean spin temperature of 100 K for interstellar hydrogen, the H I column density $N_{\text{H I}}$ to GT2318+620, measured from the absorption spectrum, is $1.5 \times 10^{21} \text{ cm}^{-2}$. From the relationship between colour excess and $N_{\text{H I}}$ derived by Kerr and Knapp⁷, this implies that the reddening $E_{B-V} = 0.3$ and extinction $A_V \approx 1$. For a distance of 3–6 kpc, the absolute visual magnitude is then $M_V = 5$ –7 mag. The absence of an optical image in the blue PSS plate indicates $B - V > 1$, or an intrinsic $(B - V)_0 \geq 0.7$, which together with the absolute magnitude suggests a late-type main-sequence dwarf.

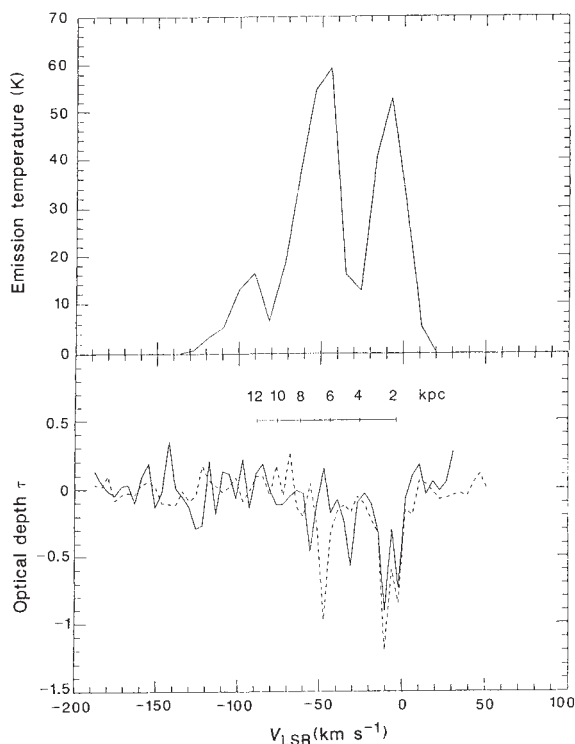


FIG. 1 Lower panel, neutral-hydrogen absorption spectra for GT2318+620 (solid line) and a nearby background double source (dashed line). Optical depth τ is dimensionless. Upper panel, spectrum of hydrogen emission in the same direction. The velocity v_{LSR} is in the frame of the local standard of rest. The spectrum of GT2318+620 lacks the absorption feature at -50 km s^{-1} , visible in the background spectrum, that is due to gas in the Perseus arm of the Galaxy. A rough kinematic distance scale is shown above the spectra.

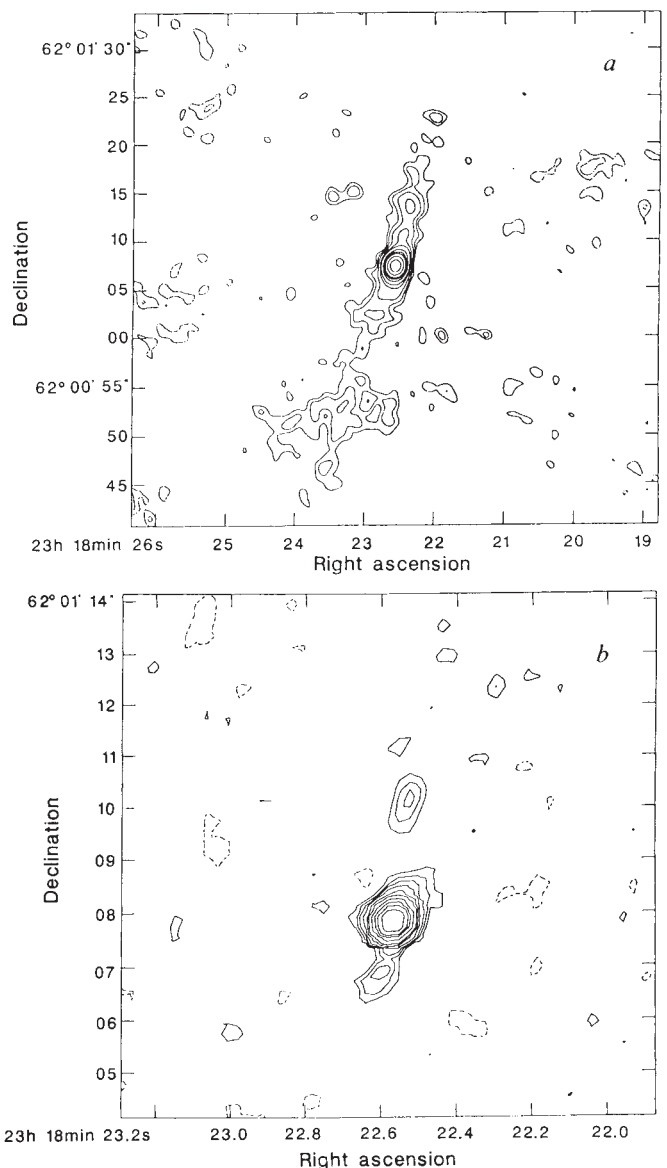


FIG. 2 a, Image of GT2318+620 at $\lambda = 20$ cm obtained on 4 November 1988 with the VLA. The angular resolution is $1.5''$. Contours are plotted at $-0.35, -0.25, 0.25, 0.35, 0.5, 0.75, 1.0, 1.5, 2.0, 2.5$ and 5.0 mJy per beam. b, Image of GT2318+620 at $\lambda = 6$ cm with angular resolution of $0.5''$. Contours are plotted at $-0.15, 0.15, 0.25, 0.35, 0.5, 0.75, 1.0, 1.5, 2.0, 2.5, 3.5$ and 5.0 mJy per beam.

The association of the UHURU X-ray source with GT2318 + 620 is less certain. The 90% probability X-ray error box, centred at $\alpha = 23^{\text{h}} 16^{\text{m}} 36^{\text{s}}$, $\delta = 61^{\circ} 48' 00''$, has an area of 0.254 square degrees. There are, however, no other obvious candidates within the error box. Moreover, it seems likely that the X-ray source is also variable. There is no catalogued Einstein X-ray source in the UHURU error box. A very weak Einstein source was found $\sim 10'$ outside the UHURU error box at $\alpha = 23^{\text{h}} 12^{\text{m}} 20^{\text{s}}$, $\delta = 61^{\circ} 35' 34''$. Unfortunately, the position of the radio source in the Einstein field centred on the UHURU position is hidden by a rib structure. It is therefore impossible to rule out the presence of a weak Einstein source at the location of GT2318 + 620.

Assuming that both the optical object on the PSS plate and the X-ray source 4U2316 + 61 are associated with GT2318 + 620, their relative luminosities provide a clue to the nature of the source. The ratio of X-ray to optical luminosity, $L_{\text{x}}/L_{\text{opt}}$, is largely independent of distance D , although strongly dependent on extinction. For $D = 3\text{--}6$ kpc, the X-ray luminosity, L_{x} lies in the range of $4 \times 10^{34}\text{--}2 \times 10^{35}$ erg s $^{-1}$ in the UHURU 2–6 keV band, yielding the ratio $L_{\text{x}}/L_{\text{opt}} \sim 100$. This value is typical of low-mass X-ray binaries, such as Sco X-1 and Cyg X-2, which have ratios of X-ray (2–11 keV) to optical luminosity of 500 and 150 respectively⁸. The optical colour of the star is consistent with a cool star, as expected for a low-mass X-ray binary, although it would be among the reddest of known low-mass X-ray binary systems. The X-ray luminosity, even at the minimum distance of ~ 3 kpc, is much higher than observed from flare stars, such as RS CVn binary systems which are more typically 10^{32} erg s $^{-1}$ (ref. 9). The X-ray luminosity is at the low end of the luminosities of X-ray binaries and at the high end of the luminosities of cataclysmic variables.

If the identifications are correct, GT2318 + 620 joins a small list of radio-emitting, low-mass X-ray binaries. Only two other X-ray binary systems, SS433 and Sco X-1, are known to have jet, or jet-like, features on a scale of an arcsecond or greater. For a distance of 4 kpc the radio luminosity of GT2318 + 620 of $\sim 2 \times 10^{20}$ erg s $^{-1}$ Hz $^{-1}$ is much higher than that of other low-mass X-ray binary systems¹⁰. The luminosity lies between that of the low-mass system Sco X-1 and the high-mass X-ray binary SS433. The bright radio jet is, perhaps, more indicative of a high-mass system. GT2318 + 620 may also be related to X-ray binary systems found in globular clusters. Recent observations have revealed weak extended radio structure associated with at least two of these systems^{11,12}. Further observations are needed to confirm the nature of GT2318 + 620. Of particular value would be an X-ray observation to confirm the identification with 4U2316 + 61, an optical spectrum to look for the presence of emission lines, infrared and optical photometry to measure more accurate colours and to search for short-period (a few hours) photometric variations caused by orbital motion, and further radio monitoring to confirm variations and to search for changes in the jet morphology similar to those seen in SS433. $\square$

Electron-doped superconductivity at 40 K in the infinite-layer compound $\text{Sr}_{1-y}\text{Nd}_y\text{CuO}_2$

M. G. Smith, A. Manthiram, J. Zhou, J. B. Goodenough & J. T. Markert*

Center for Materials Science and Engineering, ETC#5.160, University of Texas at Austin, Austin, Texas 78712-1084, USA

* Department of Physics, University of Texas at Austin, Austin, Texas 78712-1084, USA

THE known copper oxide superconductors all have intergrowth structures consisting of superconducting layers of fixed oxygen content alternating with non-superconducting oxide layers. Siegrist *et al.*¹ reported the synthesis of the 'infinite-layer' parent compound of the copper oxide superconductor—a structure comprising CuO_2 planes separated only by calcium and strontium atoms, with the composition $\text{Ca}_{0.86}\text{Sr}_{0.14}\text{CuO}_2$. Although this compound has not been made to superconduct, here we show that the isostructural compound $\text{Sr}_{1-y}\text{Nd}_y\text{CuO}_2$, in which the substitution of neodymium for strontium adds electrons to the CuO_2 planes, is superconducting at 40 K. The difference in dopant type seems to distinguish this superconductor from the superconductivity found in a multiphase sample of nominal composition $\text{Sr}_{1-y}\text{Ba}_y\text{CuO}_2$ by Takano *et al.*², in an independent study.

The superconducting layers of the copper oxide superconductors contain at least one CuO_2 sheet and have a fixed oxygen content. To induce superconductivity, the CuO_2 sheets must be oxidized (reduced) above (below) the formal oxidation state $(\text{CuO}_2)^{2-}$, but the same oxygen coordination must be maintained at all the copper atoms of a sheet. Whether the $(\text{CuO}_2)^{2-}$ sheets of an antiferromagnetic parent compound can be oxidized or reduced sufficiently to become superconducting seems to be controlled principally by the Cu–O bond length and the oxygen coordination number of the copper atoms^{3,4}. The known p-type superconductors (oxidized CuO_2 sheets) all have fivefold or sixfold coordination of the copper in the CuO_2 sheets and a lattice parameter $a \leq 3.87$ Å; the n-type T' phases have copper in fourfold coordination. But if the Cu–O bond length in the CuO_2 planes of the T' structures is too small, as in Gd_2CuO_4 ($a = 3.896$ Å), the planes do not readily accept antibonding electrons into the upper $\sigma_{x^2-y^2}^*$ band^{3,4}.

The compound $\text{Ca}_{0.86}\text{Sr}_{0.14}\text{CuO}_2$ contains fourfold-coordinated copper in an infinite sequence of CuO_2 planes separated by planes of alkaline earth atoms¹. The lattice parameter $a = 3.861$ Å is too small to allow it to be electron-doped (n-type); attempts to dope it with holes (p-type) also seem to have failed. Given the fourfold coordination of copper in this compound, our strategy was to increase the lattice parameter, and hence

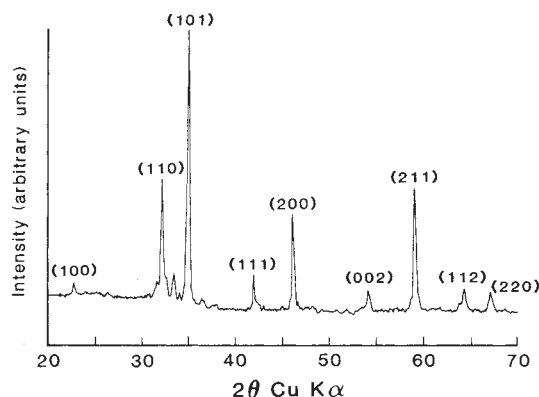


FIG. 1 X-ray powder diffraction pattern of $\text{Sr}_{0.86}\text{Nd}_{0.14}\text{CuO}_2$. The unindexed peaks are due to impurity phases.

Received 1 October; accepted 11 April 1991.

- Gregory, P. C. & Taylor, A. R. *Astr. J.* **92**, 371–411 (1986).
- Forman, W. *et al. Astrophys. J. Suppl. Ser.* **38**, 357–412 (1978).
- Hjellming, R. M. & Johnston, K. J. *Nature* **290**, 100–107 (1981).
- Weaver, H. & Williams, D. R. W. *Astr. Astrophys. Suppl.* **8**, 1–503 (1973).
- Brand, J. thesis, Sterrenwacht, Leiden (1986).
- Condon, J. J., Broderick, J. J. & Seielstad, G. A. *Astr. J.* **97**, 1064–1073 (1989).
- Kerr, F. J. & Knapp, G. R. in *Galactic Radio Astronomy* (eds Kerr, F. J. & Simonson, S. C. III) 179–182 (Reidel, Dordrecht, 1974).
- Bradt, H. V. D. & McClintok, J. E. *Ann. Rev. Astr. Astrophys.* **21**, 13–66 (1983).
- Walter, F. M. & Boyer, S. *Astrophys. J.* **245**, 671–676 (1981).
- Geldzahler, B. J. *Publ. astr. Soc. Pacif.* **99**, 1036–1038 (1987).
- Lehto, H. J., Machin, G., McHardy, I. M. & Callanan, P. *Nature* **347**, 49–51 (1990).
- Machin, G., Lehto, H. J., McHardy, I. M., Callanan, P. & Charles, P. A. *Mon. Not. R. astr. Soc.* **246**, 237–242 (1990).

ACKNOWLEDGEMENTS. This research was supported by the NSERC Canada. NRAO is operated by Associated Universities Inc., under contract from the NSF. The WSRT is operated by the Netherlands Foundation for Radio Astronomy, with the financial support of the Netherlands Organization for the Advancement of Pure Research.

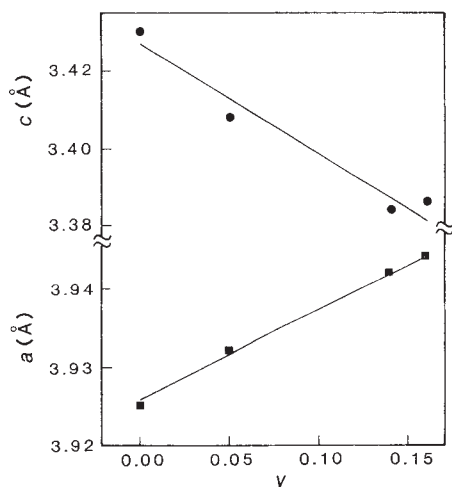


FIG. 2 Variation of lattice parameters with dopant concentration y for $\text{Sr}_{1-y}\text{Nd}_y\text{CuO}_2$.

the Cu-O bond length, by preparing an isostructural SrCuO_2 parent compound. Synthesis of this compound requires a high-pressure preparation⁵ which takes advantage of the fact that the Sr-O bond is more compressible than the Cu-O bond. Following the procedure of Takano *et al.*⁵, we obtained SrCuO_2 with lattice parameter $a=3.925$ Å at ambient pressure, which is large enough to allow reduction of the CuO_2 planes; furthermore, substitution of praseodymium or neodymium for strontium can reduce the CuO_2 planes without changing the oxygen content.

Compounds of the formula $\text{Sr}_{1-y}\text{Ln}_y\text{CuO}_2$ (where Ln is Pr or Nd) for $0 \leq y \leq 0.16$ were prepared through a nitrate route as described below. The required quantities of SrCO_3 , copper metal and Nd_2O_3 or Pr_6O_{11} were dissolved in 2M HNO_3 and the solution was evaporated to dryness on a hot plate. The resulting solid was slowly heated in air to 600 °C to decompose the nitrates, held at that temperature for 1 h, ground and re-fired at 600 °C for another 3 h. The powder was reground, made into pellets and re-fired at 925 °C for 12–16 h. The resulting product was found by X-ray powder diffraction to be a multiphase mixture without any infinite-layer phase. The pellets were finally annealed at ~1,000 °C under a hydrostatic pressure of 25 kbar for 0.5 h in a belt apparatus and quenched to room temperature before releasing the pressure to obtain $\text{Sr}_{1-y}\text{Ln}_y\text{CuO}_2$. We obtained magnetic susceptibility data with a SQUID magnetometer in a field of 10 Oe, and resistivity data using a four-probe method.

Nearly single-phase samples of $\text{Sr}_{1-y}\text{Ln}_y\text{CuO}_2$ (Ln = Pr or Nd) were obtained for $0 \leq y \leq 0.16$. X-ray powder diffraction using Cu K α radiation (Fig. 1) showed additional peaks due

to <10% unidentified impurities; no diffraction peaks corresponding to $\text{T-Ln}_{2-y}\text{Sr}_y\text{CuO}_4$, $\text{T'-Ln}_2\text{CuO}_4$ or $\text{Nd}_{2-y}\text{Sr}_{1+y}\text{Cu}_2\text{O}_6$ were observed⁶. The closer the synthetic temperature to the melting point, the lower was the amount of impurity phases. We observe a systematic variation of the lattice parameters with dopant concentration y in nearly single-phase samples (Fig. 2). As found⁷ in electron-doped $\text{Nd}_{2-y}\text{Ce}_y\text{CuO}_4$, the lattice parameter a increases and c decreases with doping. The addition of electrons into the antibonding $\sigma_{x^2-y^2}$ orbitals of the CuO_2 sheets stretches the Cu-O bonds and increases a , while the replacement of a larger Sr^{2+} by a smaller Nd^{3+} or Pr^{3+} decreases the spacing between the adjacent CuO_2 sheets and hence the lattice parameter c . Preliminary Hall data give Hall coefficient $R_H > 0$, but this could be due to the multiphase nature of the samples. The magnitude of the Hall coefficient is the sum of contributions from the electron-doped phase and other impurity phases which are probably hole-doped, and so the net Hall voltage could be positive. Unfortunately, the small sample size (~50 mg) and the presence of impurities have prevented us from making an accurate measurement of the oxygen content, which would aid in determining the sign of the carriers. But we believe that the bond-length data shown in Fig. 2 demonstrate that the CuO_2 sheets of the infinite-layer compound $\text{Sr}_{1-y}\text{Nd}_y\text{CuO}_2$ are reduced as in $\text{T'-Nd}_{2-y}\text{Ce}_y\text{CuO}_4$.

Figure 3 shows the field-cooled magnetization data for representative samples of $\text{Sr}_{1-y}\text{Nd}_y\text{CuO}_2$. Filamentary (~0.3% Meissner fraction) superconductivity was observed in the $y=0.05$ sample. Bulk superconductivity was observed for $0.14 \leq y \leq 0.16$ with onset temperatures of 34–40 K. The magnitude of the Meissner fraction varied from 6 to 24% of that expected for an ideal compact diamagnet. Both the screening and Meissner fraction increased with decreasing amount of impurity phases. The Meissner fraction was generally about half as large as the screening signal. Furthermore, no superconductivity was observed down to 2 K in undoped SrCuO_2 prepared under identical conditions. These observations demonstrate that the observed bulk superconductivity is due to the electron-doped $\text{Sr}_{1-y}\text{Nd}_y\text{CuO}_2$ and not due to any impurity phase; also, no possible impurity phases are known superconductors. The resistivity data for nominal $\text{Sr}_{0.86}\text{Nd}_{0.14}\text{CuO}_2$ (Fig. 4) show a transition temperature $T_c=40$ K. Although T_c is higher than that of any n-type $\text{T'-Ln}_{2-y}\text{M}_y\text{CuO}_4$ ($\text{M}=\text{Ce}, \text{Th}$)^{7,8}, it is not much higher, which suggests that the coupling between adjacent CuO_2 planes in n-type $\text{Sr}_{1-y}\text{Nd}_y\text{CuO}_2$ is weak.

Preliminary measurements in $\text{Sr}_{1-y}\text{Pr}_y\text{CuO}_2$ give similar results. Samples of composition $\text{Sr}_{0.85}\text{Pr}_{0.15}\text{CuO}_2$ showed a superconducting transition starting at 39 K with a Meissner fraction of ~6%. As Pr^{3+} in eightfold coordination suppresses p-type superconductivity in $\text{Y}_{1-y}\text{Pr}_y\text{Ba}_2\text{Cu}_3\text{O}_{6+x}$ (refs 9, 10), the observation of bulk superconductivity in $\text{Sr}_{1-y}\text{Pr}_y\text{CuO}_2$ provides further support for n-type superconductivity, in addition to the lattice parameters $a=3.942$ Å and $c=3.393$ Å, which are

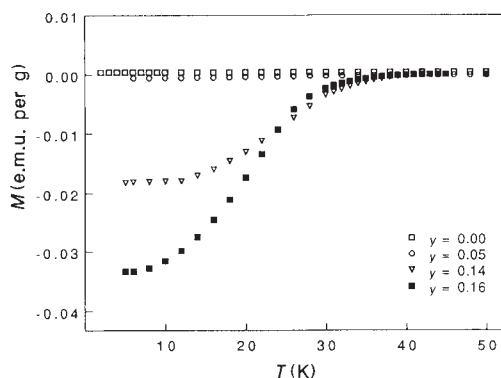


FIG. 3 Field-cooled magnetization, M , against temperature in an applied field $H=10$ Oe for $\text{Sr}_{1-y}\text{Nd}_y\text{CuO}_2$.

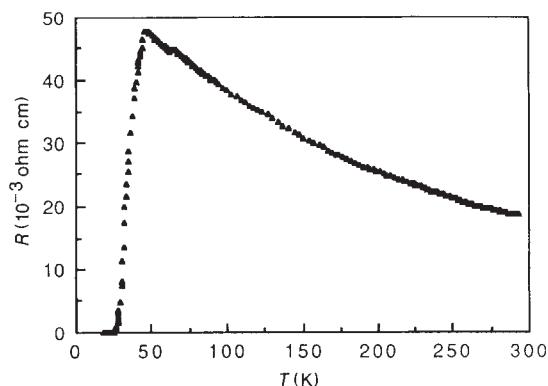


FIG. 4 Variation of resistivity with temperature for $\text{Sr}_{0.86}\text{Nd}_{0.14}\text{CuO}_2$.

increased and decreased respectively compared with their values in SrCuO_2 .

Takano *et al.*² have reported superconductivity ($T_c = 60\text{--}90\text{ K}$) in a multiphase sample obtained at 65 kbar from a starting composition $\text{Sr}_{1-y}\text{Ba}_y\text{CuO}_2$, but they did not identify which phase was superconducting. Without a Ln^{3+} ion, only p-type doping of the CuO_2 planes by means of strontium vacancies, $\text{Sr}_{1-\delta}\text{CuO}_2$, is compatible with a fixed oxygen concentration in this case.

We have demonstrated here that the in-plane Cu-O bond length is important in determining whether a CuO_2 plane can be oxidized or reduced. It would therefore be instructive to compare the lattice parameter a in the superconductor produced by Takano *et al.*² with our electron-doped superconductors.

Received 22 May; accepted 30 May 1991.

1. Siegrist, T., Zahurak, S. M., Murphy, D. W. & Roth, R. S. *Nature* **334**, 231–232 (1988).
2. Takano, M., Azuma, M., Hiroi, Z., Bando, Y. & Takeda, Y. *Physica C* (in the press).
3. Goodenough, J. B. *Supercond. Sci. Technol.* **3**, 26–37 (1990).
4. Goodenough, J. B. & Manthiram, A. *J. Solid St. Chem.* **88**, 115–139 (1990).
5. Takano, M., Takeda, Y., Okada, H., Miyamoto, M. & Kusaka, T. *Physica C* **159**, 375–378 (1989).
6. Gopalakrishnan, J., Subramanian, M. A., Torardi, C. C., Atfield, J. P. & Sleight, A. W. *Mat. Res. Bull.* **24**, 321–330 (1989).
7. Tokura, Y., Takagi, H. & Uchida, S. *Nature* **337**, 345–347 (1989).
8. Markert, J. T. & Maple, M. B. *Solid St. Commun.* **70**, 145–147 (1989).
9. Soderholm, L. *et al. Nature* **328**, 604–605 (1987).
10. Tang, X. X., Manthiram, A. & Goodenough, J. B. *Physica C* **161**, 574–580 (1989).

ACKNOWLEDGEMENTS. We thank the NSF, Texas Advanced Research Program and the R. A. Welch Foundation, Houston, Texas, for financial support.

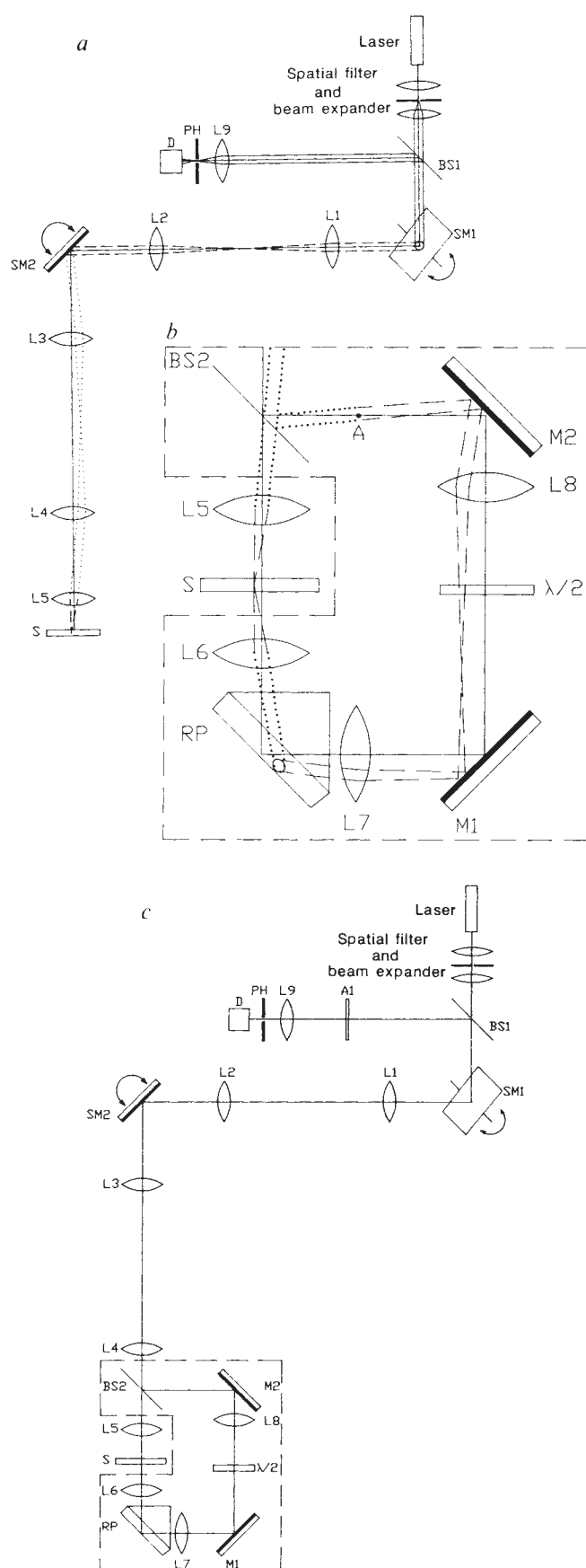


FIG. 1 Description of the optical path of the microscope. *a*, Infinity-corrected, scanning-beam confocal laser reflection microscope. *b*, Additional optical components required for imaging in transmission. *c*, Complete confocal scanning-beam reflection/transmission microscope.

A scanning confocal microscope for transmission and reflection imaging

A. E. Dixon, S. Damaskinos & M. R. Atkinson

Guelph-Waterloo Program for Graduate Work in Physics,
Waterloo Campus, University of Waterloo, Waterloo, Ontario,
Canada N2L 3G1

IN a confocal scanning laser microscope^{1,2} (CSLM), the detector pinhole is confocal with the illuminated spot on the sample, and rejects light reflected from objects that are not in the focal plane. This results in images that contain only sharp or empty areas, as opposed to non-confocal images which contain sharp and blurry areas, and allows the CSLM to perform optical sectioning. Individual optical sections can be used to produce a true three-dimensional image or can be added together to produce an extended-depth-of-focus image. For biological specimens, the intensity of the reflected-light (or fluorescence) signal decreases with increasing penetration into the sample, and self-shadowing also limits the information in the image. Moreover, many biological specimens are only weak reflectors, but produce transmission images with good contrast. The only confocal transmission images reported previously have used scanning-stage microscopes^{3,4}, although Goldstein⁵ has reported a confocal scanning-beam microscope that should in principle work in transmission. Here we describe a microscope that produces true confocal images in transmission, and also forms a reflected-light image from both the top and bottom of the specimen. The optical slices produced in transmission do not change in average intensity with depth in the specimen, and as the microscope can also examine the specimen from the bottom in reflected light, loss of data by self-shadowing is minimized. All of the original contrast modes of the CSLM (reflected light, fluorescence, differential phase contrast and so on) are also available using this new microscope.

Figure 1a is a schematic diagram of an infinity-corrected

confocal scanning-beam reflection microscope. The incoming laser beam passes through a beamsplitter (BS1) and is deflected by a scanning mirror (SM1) whose scan axis is in the plane of the diagram. Lenses L1 and L2 bring the beam back to the optic axis at SM2, which scans about an axis perpendicular to the diagram. The beam is shown above the plane of the diagram between SM1 and SM2 (shown with dashed lines) and below the plane of the diagram between SM2 and L5 (dotted lines). Lenses L3 and L4 bring the scanning beam back to the axis at the entrance pupil of the objective lens (L5), which then focuses the beam to a diffraction-limited spot at the specimen (S). Light reflected back from the specimen is collected by L5, passes back through the scan system and is partially reflected by BS1 towards the detector. Lens L9 focuses the returning light on the pinhole, and light reflected back from the position of the spot focused on the specimen passes through the pinhole and is detected. The scan mirrors are computer-controlled to raster the focal spot across the sample, and the image is stored point-by-point and displayed on a high-resolution computer screen. In general L5 is a microscope objective with a high numerical aperture (focuses a very wide cone of rays on the specimen) and a short focal length.

A practical transmission CSLM has not been developed previously because of the difficulty of descanning the transmitted beam in exact synchronism with the incoming beam. Our microscope overcomes that problem by injecting the transmitted beam

back into the original optical path in parallel with the reflected-light beam coming back from the specimen. The dashed line around the outside of Fig. 1*b* encloses the additional optical components that are necessary. L5 and S denote the microscope objective lens and sample shown at the bottom of Fig. 1*a*. The incoming laser beam at the top of Fig. 1*b* is below the plane of the diagram (dotted lines) but is moving up and to the left to cross the optic axis at the entrance pupil of the microscope objective (L5). Light transmitted through the sample is collected by another objective lens (L6). Lenses L7 and L8 perform the same function as L3 and L4, and the transmitted-light path is folded using mirrors M1 and M2 and a roof prism (RP). The roof prism also reverts the transmitted beam (reflects it from its position below the plane of the diagram (dotted lines) to above the plane of the diagram (dashed lines)). This is necessary so that the beam to be injected into the return path will cross the optic axis at A and be recombined with the reflected-light beam by the beamsplitter BS2 in exactly the right direction. (When viewed from the top of the diagram, the image of point A formed in the partially silvered mirror BS2 falls at the centre of the entrance pupil of L5.) Thus, the transmitted-light beam will pass back through the original scan system of the microscope, along with the reflected-light beam, and both signals can be detected by the same detector.

A half-wave plate ($\lambda/2$ plate) has been placed in the transmission arm, rotating the polarization of the transmitted-light beam by 90° . Thus, if an analyser is placed in front of L9 in Fig. 1*a*, either the transmitted-light signal or the reflected-light signal can be selected by rotating the analyser by 90° .

In addition to being used to inject the transmitted-light beam

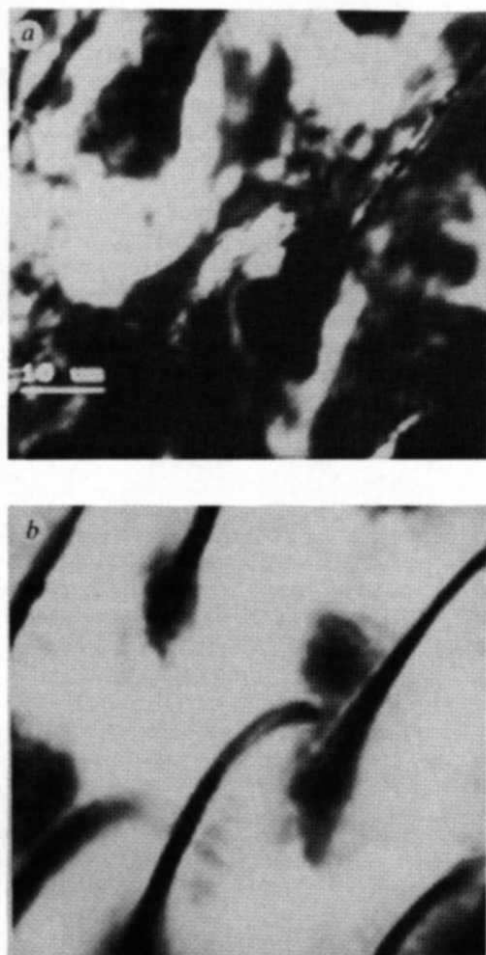


FIG. 2 Confocal images of a $60 \times 60\text{-}\mu\text{m}$ area of a fly's wing, focused on the hairs protruding above the wing. *a*, Reflection image; *b*, transmission image. All images presented here were obtained using a He-Ne laser source, and long working distance $100\times$ microscope objectives with numerical aperture 0.7. The scan rate was set to record a 512×512 image in 2 seconds. None of the images was frame averaged.

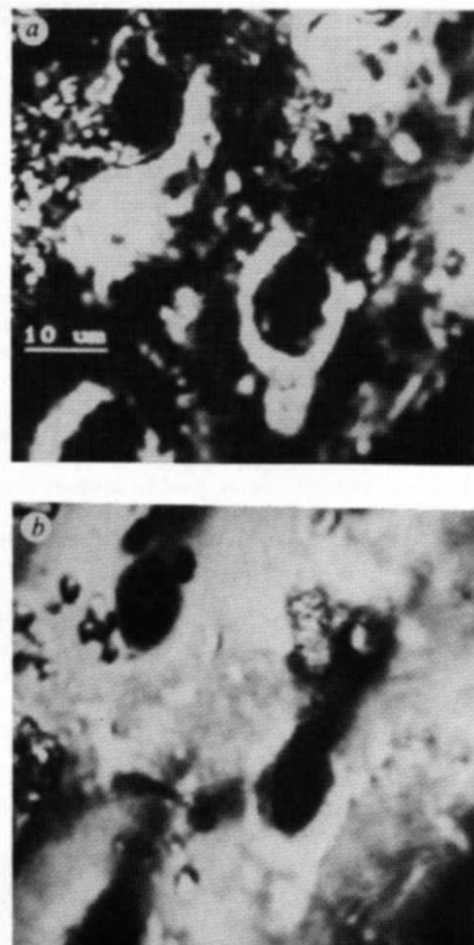


FIG. 3 Confocal images of the same $60 \times 60\text{-}\mu\text{m}$ area of a fly's wing, focused on the top surface of the wing. *a*, Reflection image; *b*, transmission image.

back into the optical path, BS2 also reflects part of the incoming laser beam towards A, where it continues around the loop to illuminate the sample from beneath. Light reflected from the bottom of the sample passes back around the loop a second time (thus passing through the $\lambda/2$ plate twice) and is injected into the returning beam at BS2 with the same polarization as the other reflected-light beam. A second transmission beam passes through the specimen from beneath, and is collected by L5. Like the other transmission beam, this one has passed through the $\lambda/2$ plate once, and is polarized perpendicular to the reflected-light beam when it reaches the detector. Note that the objective lenses L5 and L6 have high numerical aperture and short focal length, and most of the light emerging from one will be collected by the other even at the largest scan angles used in most scanning laser microscopes.

The complete transmission/reflection microscope is shown in Fig. 1c. The additions to the reflection microscope shown in Fig. 1a are the transmission scan components (enclosed by a dashed line at the bottom of the diagram) and an analyser (A1) in the detection arm of the microscope.

The wing of a common housefly was chosen as a specimen to illustrate some of the imaging modes of the microscope. Small hairs protrude from the top of the wing, and these weakly-reflecting hairs are difficult to separate from the skin of the wing in reflected light, as the skin is a very strong reflector. Figure 2a shows a single $60 \times 60\text{-}\mu\text{m}$ optical section in reflected light, focused on the hairs protruding above the surface of the wing. This measurement was obtained by blocking off the transmission arm of the microscope and is therefore identical to what would be obtained by an ordinary confocal scanning laser microscope. The weakly reflecting hairs can be seen superimposed on a strong background signal from the surface of the wing. Figure 2b shows the transmission image from the same focus position. Analyser A1 was rotated 90° to block off the reflected beam. Here the hairs have scattered and refracted light out of the transmitted beam, but the top surface of the wing is not very effective at producing contrast in the transmitted beam, which is already out of focus when it reaches the skin of the wing. For this particular specimen, there is a clear advantage in using confocal transmission to study the hairs.

Figure 3 shows optical sections when the microscope is focused on the top surface of the wing. In the reflection image (Fig. 3a) the edges of the hair follicle are bright, and this area is also bright in the transmission image (Fig. 3b). This indicates that the edges of the follicle are good reflectors, but do not absorb or scatter light out of the transmitted beam, so a large transmission intensity also results. In general, several different contrast mechanisms have contributed to the observed transmission image, and the information complements that in the reflection image.

Figure 4 illustrates the confocal nature of the transmission microscope. A confocal transmission image is shown on the left and a non-confocal transmission image on the right. The non-

confocal image was detected by mounting a collecting lens and detector beneath the sample stage. The confocal image shows a very small depth of focus ($\sim 1\text{ }\mu\text{m}$), and the confocal nature of the image is clearly shown by the rejection of signal from the out-of-focus hairs. The non-confocal picture has a much greater depth of focus, and blurry areas are evident. $\square$

Received 10 January; accepted 29 April 1991.

1. Wilson, T. & Sheppard, C. J. R. *Theory and Practice of Scanning Optical Microscopy* (Academic, London, 1984).
2. Pawley, J. (ed.) *The Handbook of Biological Confocal Microscopy* (IMR, Madison, 1989).
3. Brakenhoff, G. J. *J. Microsc.* **117**, 233–242 (1979).
4. Dixon, A. E., Damaskinos, S. & Atkinson, M. R. *Scanning* (in the press).
5. Goldstein, S. *J. Microsc.* **153**, RP1–RP2 (1989).

ACKNOWLEDGEMENTS. This work was supported by a Strategic Grant from the Natural Sciences and Engineering Research Council of Canada.

A new growth instability in colloidal crystallization

Alice P. Gast & Yiannis Monovoukas

Department of Chemical Engineering, Stanford University, Stanford, California 94305-5025, USA

SUSPENSIONS of monodisperse colloidal particles are useful model systems for studying order–disorder transitions. The bright iridescence of colloidal crystals, resulting from diffraction of visible light, is well-known^{1–5}. We have shown recently that this diffraction process depolarizes light sufficiently to reveal the orientation of the primary diffraction planes through the colours observed in a polarizing microscope^{6,7}. In the course of these investigations we have observed a new crystallization instability that can be explained, in part, by a diffusion-limited crystallization process. Although it is analogous to the mechanism of pattern formation in molecular systems, here the instability occurs for particles interacting solely by repulsive forces. We provide a qualitative explanation of the process using a classic Mullins–Sekerka⁸ stability analysis. Our observations suggest that colloidal suspensions can provide a useful model not only for the study of crystallization thermodynamics but also for investigations of ordering dynamics.

Crystals form in aqueous suspensions of charged particles when critical volume fraction, ϕ_c , is reached, or when the ionic strength is reduced to the point where particles interact over a range such that they become close packed^{5,9}. One can describe this ordering by treating the particles as effective hard spheres. The screened coulomb repulsive energy of magnitude α and range κ , $u(r)/kT = \alpha \exp(-\kappa r)/\kappa r$, where r is radius, can be approximated with a hard sphere diameter, d_{eff} , defined by requiring the chemical potential for the effective hard-sphere system to equal that of the real ensemble¹⁰. Hard spheres order at volume fractions $\phi_h = 0.50$ and volume-filling crystals of hard particles first melt at $\phi_h = 0.55$ (ref. 11).

Our experiments involve highly charged polystyrene particles, of radius $a = 67\text{ nm}$, suspended in thin capillary cuvettes, $100\text{--}400\text{-}\mu\text{m}$ thick, at an ionic strength of $3 \times 10^{-5}\text{ M}$ added salt and at volume fractions spanning the solid–liquid coexistence regime^{2–4}. The magnitude of the potential, α , is $\sim 10^6$, and the range is comparable to the particle size, $\alpha\kappa = 1.76$. Under these conditions the particles freeze at a volume fraction $\phi_f = 0.016$, and the upper limit of liquid–solid coexistence is $\phi_m = 0.067$. The effective hard-sphere diameter varies from $d_{\text{eff}}/2a = 3.24$ at the freezing point to $d_{\text{eff}}/2a = 2.06$ at the melting limit, making our effective hard spheres freeze and melt at $\phi_h = 0.50$ and $\phi_h = 0.58$ respectively.

In Figs 1 and 2 we show the typical crystallization morphology, viewed through crossed polarizers in a Zeiss Axioplan

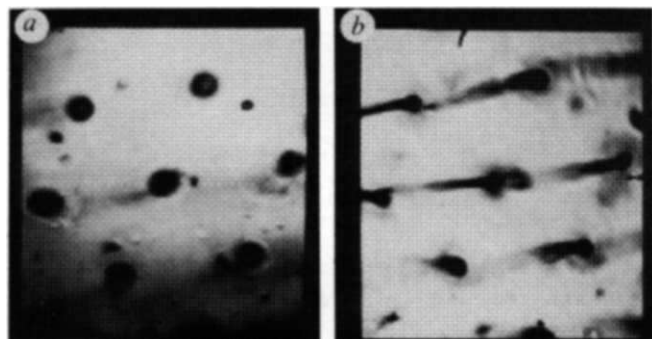


FIG. 4 a, Confocal transmission image; b, non-confocal transmission image, focused on the top surface of the wing.

microscope for a suspension at $\phi = 0.040$. At this concentration the suspension exhibits many nearly close-packed crystallites (Fig. 1), with the crystal colours indicating the lattice spacing and orientation relative to the incident light⁷. The striations visible in the crystallites are due to random switching between face-centred-cubic twin crystal structures⁶. Unstable growth in the form of fingers appears at the edges of such clusters of crystallites (Fig. 2). These fingers, about 70–150 μm in width, appear only where growing into a disordered region. This is demonstrated strikingly in Fig. 3 where a suspension of $\phi = 0.032$ produces highly fingered crystallites nucleated at the capillary walls; here, the characteristic width of a finger is 170 μm . As these crystallite fingers are only visible when growth occurs into a region of disorder, and their size increases as we approach the freezing concentration, we suggest that their formation results from the diffusion-limited transport of particles to the solid-liquid interface.

When diffusion of particles to the crystal surface becomes the rate-limiting process in crystal growth, instabilities such as those observed in molecular systems can occur. A simple analysis of these instabilities was originally developed by Mullins and Sekerka⁸. Here a plane crystal front propagates through a diffusion field¹², and perturbations to this front persist on a length scale $\lambda = 2\pi(l d/2)^{1/2}$, where $l = 2D/v$ is the diffusion length, v is the interface velocity and d is the capillary length

or interfacial thickness. The capillary length is defined by

$$d = \frac{\gamma}{\Delta\phi^2(\partial\mu/\partial\phi)}$$

where γ is the interfacial tension, $\Delta\phi$ is the miscibility gap, $\phi - \phi_c$, and μ is the chemical potential. Purely repulsive systems have interfacial tensions $\gamma a^2/kT$ of order unity¹³, making the interfacial thickness ~ 5 –10 interparticle spacings, or ~ 15 –30 a (1–2 μm for our system).

We approximate the instability wavelength for diffusion-limited crystallization in repulsive colloidal particles by determining the mutual diffusion coefficient, $D(\phi) = D_0[1 + (2A_2 + K_2)\phi]$ (ref. 9) and interfacial velocity $v = D_0\phi^{1/3}\{1 - e^{-\Delta\mu/kT}\}/a$ (ref. 14) where A_2 and K_2 are the second virial coefficient and sedimentation coefficient respectively, and D_0 is the diffusion coefficient for an isolated particle. Although a more accurate analysis of $D(\phi)$ has recently become available¹⁵, it is not warranted in the simple mechanistic analysis presented here. Calculating the approximate diffusion coefficient for our effective hard spheres from $A_2 = d_{\text{eff}}^3/2$ and $K_2 = 1 - (3d_{\text{eff}}^2/2) - (15/4d_{\text{eff}}) + (9/8d_{\text{eff}})^3 - O(d_{\text{eff}}^{-5})$ (ref. 9), we obtain $D(\phi) \approx D_0[1 + 200\phi]$. The driving force, $\Delta\mu$, is approximated by $\Delta\mu/kT \approx C\{\phi/0.016 - 1\}$ where C is a constant, equal to 7.3 for a similar suspension¹⁴.

This analysis leads to a characteristic instability wavelength

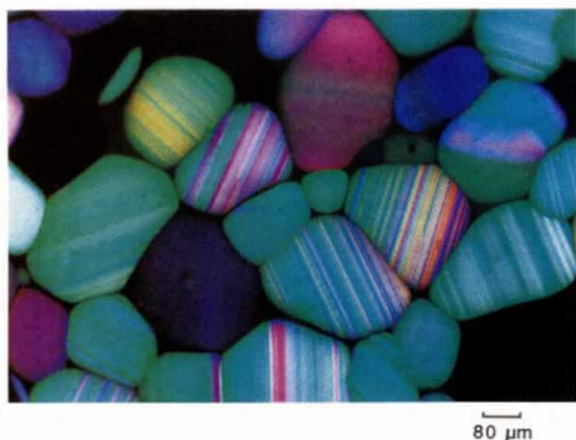


FIG. 1 Charged polystyrene particles 67 nm in radius suspended in a 200- μm glass capillary at a volume fraction $\phi = 0.04$ and observed between crossed polarizers.

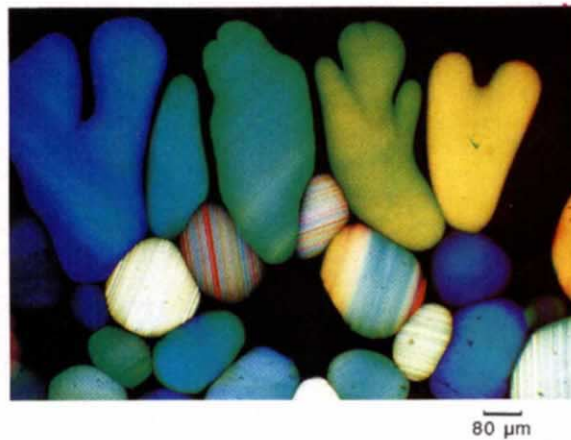


FIG. 2 The same suspension as in Fig. 1 observed at the edge of a group of crystallites.



FIG. 3 Polystyrene suspension at $\phi = 0.032$. The width of the striated crystallite is 230 μm at the bright white band.



FIG. 4 Crystallites seeded with ion-exchange resin at lower edge. Here $\phi \approx 0.02$.

ranging from 150 μm very near the freezing volume fraction to tens of micrometres within the coexistence regime. These predictions are consistent with our observations; the size of the instability is the correct order of magnitude and its concentration dependence is appropriate. We cannot observe fingers at higher concentrations because of the limited time and space for growth imposed by closely packed crystals, but we tested the hypothesis of diffusion-limited crystal growth further by seeding crystallization with ion-exchange particles. Ion-exchange particles rapidly remove ions from the suspension, reducing the concentration into the solid-liquid coexistence region. This enhances the driving force for crystallization by increasing $\Delta\mu$, thus speeding up the growth of the crystal interface and shortening the diffusion length l . This is evident in Fig. 4 where such crystals display smaller fingers, $\sim 50\ \mu\text{m}$ wide, extending over large distances within the crystal. Clearly, a more sophisticated analysis is needed to describe these complex shapes in more detail.

We have observed a new crystallization instability occurring in purely repulsive colloidal particles. Our quasi-stationary linear stability analysis supports the idea of diffusion-limited growth as the instability mechanism, but the approximations about the transport and interfacial properties of colloidal crystals preclude a more precise evaluation. A more detailed study should provide useful insight into the solid-liquid interfacial behaviour of systems of repulsive spheres. $\square$

Received 5 March; accepted 29 April 1991.

1. Pieranski, P. *Contemp. Phys.* **24**, 25–73 (1983).
2. Pusey, P. N. & van Megen, W. *Nature* **320**, 340–342 (1986).
3. Kose, A. & Hachisu, S. *J. Colloid Interface Sci.* **46**, 460–469 (1974).
4. Hiltner, A. P., Papir, Y. S. & Krieger, I. M. *J. phys. Chem.* **75**, 1881–1886 (1971).
5. Monovoukas, Y. & Gast, A. P. *J. Colloid Interface Sci.* **128**, 533–548 (1989).
6. Monovoukas, Y. & Gast, A. P. *Langmuir* **7**, 460–468 (1991).
7. Monovoukas, Y., Fuller, G. G. & Gast, A. P. *J. chem. Phys.* **93**, 8294–8299 (1991).
8. Mullins, W. W. & Sekerka, R. F. *J. appl. Phys.* **35**, 444–451 (1964).
9. Russel, W. B., Schowalter, W. R. & Saville, D. A. *Colloidal Dispersions* (Cambridge University Press, 1989).
10. Russel, W. B. *Phase Transitions* **21**, 127–137 (1990).
11. Alder, B. J., Hoover, W. G. & Young, D. A. *J. chem. Phys.* **49**, 3688–3696 (1968).
12. Langer, J. S. *Rev. mod. Phys.* **52**, 1–28 (1980).
13. Oxtoby, D. W. & McMullen, W. E. *Phys. Chem. Liq.* **18**, 97–105 (1988).
14. Austen, D. J. W., Clark, N. A., Cotter, L. K. & Ackerson, B. J. *Phys. Rev. Lett.* **57**, 1733–1736 (1986).
15. Genz, U. & Klein, R. *Physica A* **171**, 26–42 (1991).

ACKNOWLEDGEMENTS. We thank G. M. Homsy for his comments.

Intercalation of copper metal clusters in montmorillonite

P. B. Malla, P. Ravindranathan, S. Komarneni & R. Roy

Materials Research Laboratory, The Pennsylvania State University, University Park, Pennsylvania 16802, USA

EXPANDABLE layer silicates such as montmorillonite can be converted to efficient heterogeneous catalysts by introducing catalytically active sites or guest species between the layers or on the external surfaces. Attempts to produce intercalated zero-valent transition-metal particles in layer silicates, by hydrogen reduction for example, have, however, failed: the layers tend to collapse¹, sometimes followed by deposition of metal particles on the external surfaces^{2–4}. Here we describe the successful intercalation of copper metal clusters of 4–5 Å in montmorillonite by *in situ* reduction of Cu^{2+} ions using ethylene glycol. These metal-cluster intercalates were stable up to at least 500 °C. The clusters prop the silicate layers apart, much as metal oxides do in pillared clays⁵, and may thus be able to introduce unique catalytic product selectivity through a molecular sieving effect similar to that in cluster-loaded zeolites. As metal clusters of these dimensions behave very differently from the bulk metal⁶, intercalates of this sort may prove to be versatile catalysts.

Fine metal clusters supported in porous solids, such as zeolites, possess unique catalytic properties, molecular selectivity and polyfunctional activity^{6–8}. They are widely used in petroleum refining, and in the chemical and petrochemical industries. The Cu^{2+} -exchanged montmorillonite can facilitate unusual conversions of organic molecules⁹. Impregnation of $\text{Cu}(\text{NO}_3)_2$ in zirconium-pillared montmorillonite activates the conversion of CO/H_2 into hydrocarbons¹⁰. This process, however, involves two steps: preparation of pillared clays, followed by impregnation. Our objective, therefore, was to develop a single process which could produce both catalytically active sites and propping apart of layers (to produce a molecular sieve for shape and size selectivity) in montmorillonite.

We used a natural Na-montmorillonite¹¹ from Tsukinuno district, Japan, having cation-exchange capacity of 100 milliequivalent (meq) per 100 g and the structural formula of $(\text{Na}_{0.35}\text{K}_{0.01}\text{Ca}_{0.02})(\text{Si}_{3.89}\text{Al}_{0.11})(\text{Al}_{1.6}\text{Fe}_{0.08}\text{Mg}_{0.32})\text{O}_{10}(\text{OH})_2 \cdot n\text{H}_2\text{O}$. This mineral was ion-exchanged with Cu^{2+} by repeated leaching with an excess of 0.5M $\text{CuCl}_2 \cdot 2\text{H}_2\text{O}$ over a period of 24 hours at room temperature followed by washing with deionized water to remove the excess salts.

A sample (1 g) of Cu^{2+} -exchanged montmorillonite was suspended in 100 ml of ethylene glycol in a double-necked round-bottomed flask and refluxed at 195 °C for 6 hr in an argon environment^{12,13}. The solid product was washed with methanol and separated by centrifugation, then stored in a vacuum desiccator before further characterization.

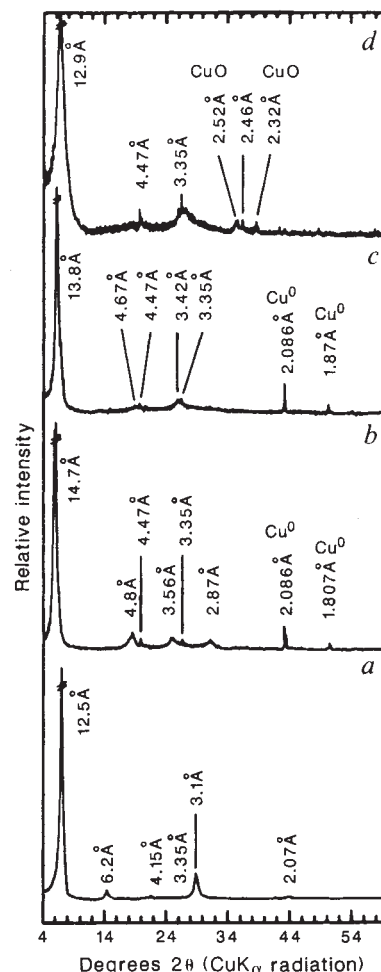


FIG. 1 X-ray powder diffraction patterns of montmorillonites: a, Cu^{2+} -exchanged; b, Cu^0 -intercalated by reducing in ethylene glycol followed by evacuation at 200 °C; c, Cu^0 -intercalated and heated in hydrogen at 400 °C for four hours; d, Cu^0 -intercalated and heated in air at 400 °C for 4 h.

Measurements of the degree of ion exchange by repeated leaching with 0.5 M $\text{MgCl}_2 \cdot 6\text{H}_2\text{O}$ (back ion-exchange method)¹⁴ indicated that the Cu^{2+} -exchanged montmorillonite and the reduced sample possessed 110 and 33 meq per 100 g (for vacuum-dried samples) of exchangeable Cu^{2+} ions, respectively. These results correspond to the reduction of 70% of the Cu^{2+} ions to Cu^0 assuming that no metal clusters were reoxidized to Cu^{2+} ions during the ion-exchange experiment.

X-ray powder diffraction patterns of Cu^{2+} -exchanged and Cu^0 -intercalated montmorillonites are compared in Fig. 1. The reduced sample, which was repeatedly washed in methanol, had a d -spacing of 15.6 Å (X-ray pattern is not shown in figure). To avoid any contribution of intercalated ethylene glycol molecules to d , the reduced sample was evacuated at 200 °C which gave $d = 14.7$ Å and a sharper peak (Fig. 1*b-d*) than for the as-prepared sample. This d -value indicated the presence of copper metal clusters 5.1 Å in diameter between layers after subtracting the thickness of the basal aluminosilicate layer, 9.6 Å. The intercalation of Cu^0 in montmorillonite can also be deduced from the disappearance of $d(002)$ peaks (Fig. 1*b-d*) as a result of heavy scattering from these intercalated Cu^0 clusters. In addition to the intercalation of copper metal clusters, some copper metal particles were deposited on the external surfaces, as indicated by the peaks at 2.086 Å (111) and 1.807 Å (200). Transmission electron microscopy also provided evidence for the intercalation of Cu^0 clusters (Fig. 2). When samples in which no Cu^0 could initially be observed were exposed to the electron beam for ≥ 5 min, the beam damaged the structure of montmorillonite with concomitant crystallization of ~ 10 -nm particles. These particles were identified as Cu^0 by convergent beam diffraction and are thought to have been grown by coalesc-

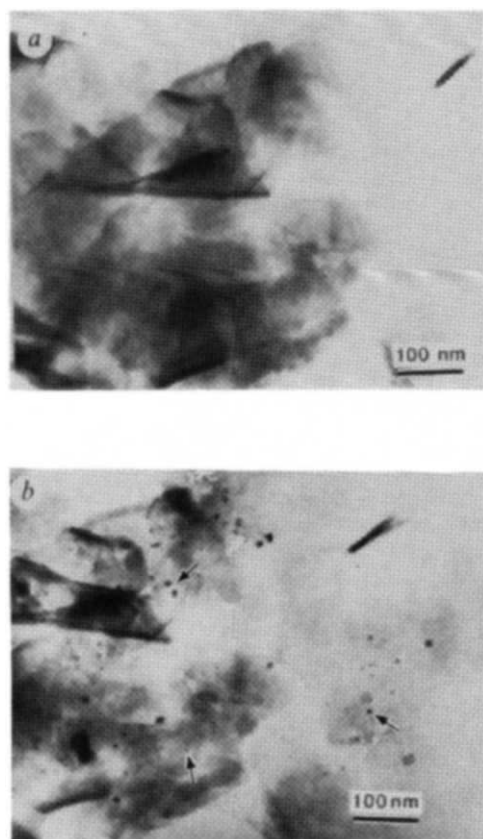


FIG. 2 Transmission electron micrographs (bright-field images) of reduced montmorillonite: *a*, after 1 min; *b*, after 5 min exposure to the electron beam. Note the appearance of numerous 3–15-nm Cu^0 crystallites (a few indicated by arrows) in (*b*).

ing the intercalated 4–5 Å Cu^0 clusters after the host matrix was rendered amorphous.

The thermal treatment of reduced samples at 400 and 500 °C in hydrogen gave a basal spacing of 13.8 Å (Fig. 1*c*) and 12.5 Å, respectively, indicating that these metal cluster intercalates have high thermal stability. Samples heated in air at 300°, 400° (Fig. 1*d*) and 500 °C had d spacings of 13.8, 12.9 and 12.2 Å respectively, which indicated that these metal intercalates are thermally stable even in air at least up to 500 °C. The metal particles deposited on the external surfaces were, however, completely converted to CuO at ≥ 400 °C (Fig. 1*d*). But one may argue that the thermal treatment at 400 or 500 °C in air converts the Cu^0 clusters to CuO in the interlayers. Earlier works on basic Cu-acetate¹¹ and montmorillonites interlayered with other transition metal hydroxides^{2–4} indicate that CuO , NiO , CoO and FeO are not stable in the interlayers above 300 °C and are always expelled onto the external surfaces, collapsing the montmorillonite layers to 9.6 Å. In contrast to the sample reduced by ethylene glycol, the Cu^{2+} -exchanged sample reduced by hydrogen collapsed to 9.8 Å at temperatures as low as 200 °C, indicating that Cu^{2+} ions or Cu^0 had migrated to the hexagonal and/or octahedral holes^{1,15}. The Cu^{2+} -exchanged sample, which expanded to 16.9 Å when exposed to ethylene glycol vapour before reduction, also collapsed to 9.6 Å after heating at ≥ 200 °C. This rules out the possible contribution of fragments of ethylene glycol in the formation of the 14.7 Å phase (Fig. 1*b*), as the reduction was followed by evacuation at 200 °C.

The pore structure of montmorillonite with intercalated copper metal clusters was probed by measuring water adsorption isotherms at 25 °C (Fig. 3) and nitrogen surface area. The water adsorption isotherms of a Cu^{2+} -exchanged sample evacuated at room temperature showed adsorption steps in accordance with the hydration of cations (Fig. 3*a*). The adsorption was severely reduced when the sample was evacuated at 200 °C (Fig. 3*d*), corroborating the powder-diffraction evidence for collapsed layers. The reduced samples intercalated with metal clusters and evacuated at 200 °C exhibited initial enhancement in adsorption, indicating their microporous nature, followed by increases at high relative pressure as a result of condensation of water in macropores and on external surfaces (Fig. 3*b*)¹⁶. The water adsorption was reduced only slightly when evacuated at 400 °C (Fig. 3*c*). The surface areas estimated from BET (Brunauer, Emmett and Teller) monolayer capacity of water ($P/P_0 = 0.05$ – 0.35) were 224 and 179 $\text{m}^2 \text{g}^{-1}$ for samples evacuated at 200 and 400 °C, respectively. The microporous nature of the

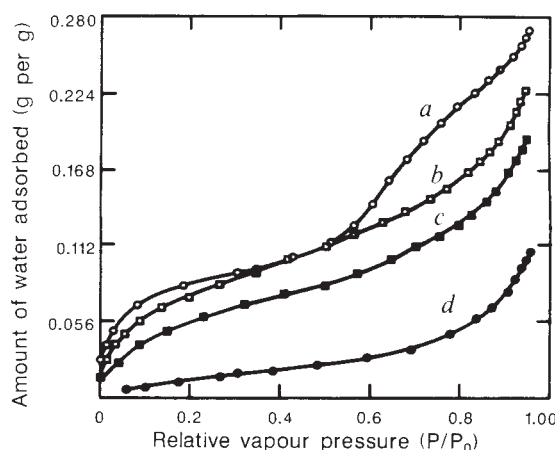


FIG. 3 Water-adsorption isotherms of montmorillonites: *a*, Cu^{2+} -exchanged and degassed at room temperature (open circles); *b*, Cu^0 -intercalated and degassed at 200 °C (open squares); *c*, Cu^0 -intercalated and degassed at 400 °C (closed squares); *d*, Cu^{2+} -exchanged and degassed at 200 °C (closed circles).

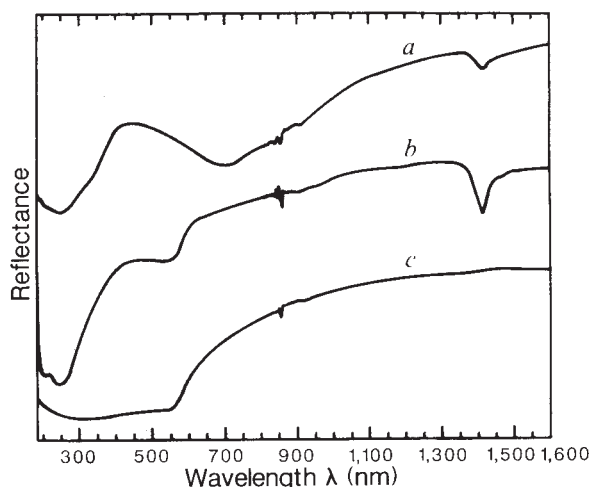


FIG. 4 Diffuse reflectance spectra of montmorillonites: *a*, Cu^{2+} -exchanged; *b*, Cu^0 -intercalated; *c*, Cu^0 prepared from hydroxy copper acetate (without support) by reducing in ethylene glycol.

reduced sample was also apparent from the high surface area of nitrogen ($117 \text{ m}^2 \text{ g}^{-1}$) compared with the unreduced sample ($30 \text{ m}^2 \text{ g}^{-1}$). Although this surface area is smaller than that of alumina-pillared clay ($200\text{--}400 \text{ m}^2 \text{ g}^{-1}$)^{17,18} which exhibits slit widths of 8–9 Å, both the water isotherms and the nitrogen surface area of montmorillonite intercalated with copper metal clusters are comparable to those of alkylammonium-pillared montmorillonites ($100\text{--}200 \text{ m}^2 \text{ g}^{-1}$) having ≤ 4.5 Å free distances between layers¹⁹.

Diffuse reflectance spectra of various samples are presented in Fig. 4. The absorption band of Cu^{2+} -exchanged montmorillonite at $\sim 700 \text{ nm}$ (Fig. 4*a*) can be attributed to a $d\text{--}d$ transition of the tetragonal square planar $\text{Cu}(\text{H}_2\text{O})_4^{2+}$ complex²⁰. The band at 250 nm indicates charge transfer from oxygen to isolated Cu^{2+} ions²¹. The reduced sample did not show any sharp absorption band in the visible region, but the absorption edge at $\sim 550\text{--}570 \text{ nm}$ (Fig. 4*b*) is close to the absorption band exhibited by copper sols at 575 nm²². In the ultraviolet region, the charge-transfer band of the reduced sample is sharper than in the unreduced sample and persists even after reduction in ethylene glycol, indicating that there is charge transfer between these intercalated metal clusters and the basal oxygen atoms. The charge transfer mechanism may have stabilized the metal clusters in the interlayers. The band at 1,410 nm (Fig. 4*a* and *b*) can be ascribed to the overtones caused by vibrations of lattice hydroxyl groups²⁰. The spectrum of pure copper metal particles prepared by reducing hydroxy copper acetate (without support) by ethylene glycol is given in Fig. 4*c* for comparison.

The possible mechanism by which ethylene glycol reduces metal oxide or hydroxide in a homogeneous system not involving support has been described by Fievet¹³ as nucleation and growth, and we believe that a similar mechanism is operative in systems involving support. In Cu^{2+} -exchanged montmorillonite, ethylene glycol, being a polar solvent, has easy access to the interlayers where it reduces individual Cu^{2+} ions. Each reduced atom then probably acts as a nucleus for other reduced atoms to grow around it, and the nuclei may coalesce to form 4–5 Å clusters. Further growth of metal clusters beyond this size may have been suppressed by steric hindrance offered by the montmorillonite layers. In the process, some copper metal clusters or nuclei were expelled to the external surfaces where they could grow freely to diameters of $\leq 0.5 \mu\text{m}$. We hope that these results will improve our fundamental understanding of metal clusters in a constrained geometry and applications related to catalysis and molecular sieving. □

Received 27 December 1990; accepted 16 April 1991.

1. Patel, M. *Clays Clay Miner.* **30**, 397–399 (1982).
2. Ohtsuka, K. *et al. Bull. Chem. Soc. Japan* **60**, 871–876 (1987).
3. Ohtsuka, K., Koga, J., Suda, M., Ono, M. & Takahashi, M. *Bull. Chem. Soc. Japan* **60**, 2843–2847 (1987).
4. Ohtsuka, K., Suda, M. & Ono, M. *Bull. Chem. Soc. Japan* **61**, 815–820 (1988).
5. Pinnavaia, T. J. *Science* **220**, 365–371.
6. Nicolaides, C. P. & Scurrall, M. S. in *Keynotes in Energy-Related Catalysis* Ch. 6 (Elsevier, New York, 1988).
7. Minachev, W. M. & Isakov, Ya. I. in *Zeolite Chemistry and Catalysis* (Am. chem. Soc. monogr. 171) 552–611 (1976).
8. Jacobs, P. A. in *Metal Clusters in Catalysis* Ch. 8 (Elsevier, New York, 1986).
9. Bartley, G. J. J. & Burch, R. *Appl. Catal.* **28**, 209–221 (1986).
10. Thomas, J. M. in *Intercalation Chemistry* Ch. 3 (Academic, London, 1982).
11. Yamanaka, S., Numata, K. & Hattori, M. in *Proc. 8th Int. Clay Conf.* (eds Schultz, L. G., van Olphen, H. & Mumpton, F. A.) 273–276 (Clay Minerals Society, Bloomington, Indiana, 1987).
12. Ravindranathan, P., Malla, P. B., Komarneni, S. & Roy, R. *Catal. Lett.* **6**, 401–408 (1990).
13. Fievet, F., Lagier, J. P., Blin, B., Beaudou, B. & Figlarz, M. *Solid St. Ionics* **32/33**, 198–205 (1989).
14. Couglan, B. & Keane, M. J. *Catal.* **123**, 364–374 (1990).
15. Hofmann, V. U. & Klemen, R. Z. *anorg. Chem.* **262**, 95–99 (1950).
16. Malla, P. B. & Komarneni, S. *Clays Clay Miner.* **38**, 363–372 (1990).
17. Pinnavaia, T. J., Tzou, M.-S., Landau, S. D. & Raythatha, R. H. *J. molec. Catal.* **27**, 195–212 (1984).
18. Ocelli, M. L. in *Proc. 8th Int. Clay Conf.* (eds Schultz, L. G., van Olphen, H. & Mumpton, F. A.) 319–323 (Clay Minerals Society, Bloomington, Indiana, 1987).
19. Barrer, R. M. *Zeolite and Clay Minerals as Sorbents and Molecules Sieves* Ch. 8 (Academic, London, 1978).
20. Velge, F., Schoonheydt, R. A. & Uytterhoeven, J. B. *Clays Clay Miner.* **25**, 375–380 (1977).
21. Borskov, G. in *Proc. 6th Int. Congress Catal.* (eds Bond, G. C., Wells, P. B. & Tompkins, F. C.) 204–215 (Chemical Society, London, 1976).
22. Curtis, A. C. *et al. Angew. Chem. Int. Ed. Engl.* **27**, 1530–1533 (1988).

ACKNOWLEDGEMENTS. We thank E. Breval for taking the transmission electron micrographs. This research was supported by the US Dept of Energy.

Evidence for a 50% increase in H_2O_2 over the past 200 years from a Greenland ice core

Andreas Sigg & Albrecht Neftel

Physics Institute, University of Bern, Sidlerstrasse 5, CH-3012 Bern, Switzerland

HYDROGEN peroxide has attracted increasing attention from atmospheric chemists over the past decade, because in the gas phase it acts as a reservoir species of OH radicals^{1,2}, and in the aqueous phase it plays a key role in the oxidation of SO_2 to H_2SO_4 in clouds³. It is also known to have an adverse effect on trees and plants⁴. In 1984, hydrogen peroxide was identified as one of the dominant trace species in polar ice⁵, introducing the possibility of constructing a record of atmospheric hydrogen peroxide concentrations from ice-core data. Here we present such a record for the past 700 years from Summit, Central Greenland. We find that hydrogen peroxide concentrations have increased by 50% over the past 200 years, with most of the increase occurring in the past 20 years, indicating that human activities may be responsible for such a dramatic change.

Ice cores were taken during the Eurocore drilling operation at Summit, Central Greenland ($72^\circ 34' \text{N}$, $37^\circ 38' \text{W}$). Hydrogen peroxide was measured by an enzymatic-fluorimetric method^{6,7}. (In this method, fluorescent dimers of 4-ethylphenol are formed by reaction with H_2O_2 and peroxidase.) A cross-section of 1 cm^2 of the core was needed for the analysis, which has a resolution of 3–5 cm in the upper firn part of the core. At depths below 30 m, we used a continuous melting and measuring technique instead, which allowed us to measure H_2O_2 with a depth resolution of better than 1 cm. We checked periodically for interference of organic peroxides by introducing a MnO_2 catalyst column, which selectively removes hydrogen peroxide into the sample line⁸. No significant contribution of organic peroxides was found.

The analysis of the Summit ice core yielded very high H_2O_2 concentrations of up to $13 \mu\text{M}$. In about the top 15 m, there is a strong seasonal cycle, which disappears completely during firnification. Figure 1 displays the firn part (0–70 m) of the

continuous H_2O_2 record. The smoothing is a consequence of evaporation, gas-phase diffusion and recondensation processes taking place in the firn. The behaviour of H_2O_2 is similar to that of H_2O , whose redistribution during firnification is shown by its ^{18}O smoothing. Hydrogen peroxide is smoothed more than water, as a significant part of the isotope seasonal signal in water survives firnification at Summit (S. J. Johnsen, personal communication). Figure 2 shows the annual mean concentration of H_2O_2 in the top 178 m of the Summit core, covering ~ 700 years. Only the top 20 m of the Summit core could be dated by counting H_2O_2 cycles. Below this, the dating is based on $\delta^{18}\text{O}$ ($[(^{18}\text{O}/^{16}\text{O})_{\text{sample}}/(^{18}\text{O}/^{16}\text{O})_{\text{SMOW}}]-1$) and electrical conductivity measurements (H. B. Clausen, personal communication). The interannual variations in the annual mean concentration are partly a result of variable precipitation throughout the year and wind drift of surface snow. To remove this high-frequency noise and to discern systematic trends, we applied a cosine-shaped numerical filter of width 40 years (solid curve in Fig. 2). The period 1300–1800 is characterized by little change, in the average H_2O_2 concentration. Between 1800 and 1950, an increase of $\sim 20\%$ is observed, followed by a local minimum around 1960. Between 1970 and 1989, H_2O_2 concentrations increase markedly to $\sim 50\%$ above pre-industrial values. For comparison, we present a 200-year H_2O_2 record from Dye 3, South Greenland ($65^\circ 11' \text{N}$, $43^\circ 49' \text{W}$), which was measured in a shallow core drilled in 1986. This record was extended at the top by *in situ* measurements in a second shallow core drilled in 1988 at the same location. At Dye 3, the seasonal H_2O_2 cycle survives firnification because of the higher accumulation rate and can be used for dating. In Fig. 3, the two Greenland records are displayed. The record for Dye 3 is not of the same quality as the Summit record. At Dye 3, surface melting is very common during summer. Melt layers can strongly influence the original H_2O_2 concentration⁹, introducing additional noise to the record. In addition, the stratigraphy of the Dye 3 core after 1960 may have been disturbed by activities of the nearby Dye 3 station, which was built in the early 1960s. At Summit, melt layers are very rare, and there was no local human activity before 1989. Nevertheless, the two records exhibit some similar features, such as a local H_2O_2 minimum around 1960 and an increase since 1970. Finally, we should mention that earlier H_2O_2 measure-

ments⁹ on an ice core from Siple Station, West Antarctica, showed no clear trend between 1900 and 1980.

The firn part of the Summit core covers the past 200 years, exactly where the H_2O_2 increase is observed. To discuss whether this increase reflects an atmospheric change or whether it is only an artefact of physical or chemical loss processes in the firn, we must describe the effect of firnification in the H_2O_2 distribution in a little more detail. The fact that the seasonal H_2O_2 signal disappears in a depth of only 35 m indicates that smoothing by firnification is straight in the upper part of the firn. Below the depth at which pores close off (70 m at Summit), a further smoothing is only possible by solid-phase diffusion in the ice, which is orders of magnitudes slower than gas-phase diffusion. According to a model by Johnsen¹⁰, the smoothing effect of firnification for water isotopes can be described mathematically by a gaussian low-pass filter. A characteristic diffusion length, L , which is the root-mean-square vertical displacement of the molecules from their original location, determines the width of the filter. A signal of wavelength λ (for example a seasonal cycle) is attenuated by the factor $R = \exp(-2\pi^2 L^2/\lambda^2)$ during firnification. The diffusion length and wavelength are expressed in cm of ice. An estimate of $L \approx 8$ cm was obtained for water isotopes by comparing the amplitude attenuation of the seasonal signal in cores with different accumulation rates¹⁰. A previous estimate for the diffusion length of H_2O_2 ($L \approx 8$ cm) was based on the H_2O_2 record from Siple, Antarctica (accumulation rate $A = 0.5 \text{ m yr}^{-1}$)⁹. The new Summit data ($A = 0.23 \text{ m yr}^{-1}$) and H_2O_2 measurements on a shallow core from Crête (Central Greenland, $A = 0.30 \text{ m yr}^{-1}$) show that a diffusion length of $L = 12$ cm is a better estimate for hydrogen peroxide. This means that during firnification the displacement of the H_2O_2 molecules from their original position is limited to distances much less than one metre. A possible outgassing effect is therefore only expected in the first metre of the firn. Similarly, a possible chemical loss of H_2O_2 in firn air is most probably restricted to the uppermost layers, where ultraviolet light and short-lived OH radicals can penetrate. In the dark open-pore space of the firn, chemical decomposition of hydrogen peroxide is therefore unlikely. A second, independent argument can be used: Dye 3 has an accumulation rate nearly 2.5 times as high as that of Summit, leading to a different depth–age relationship. In addition, the higher temperatures at Dye 3 lead to a faster firnification¹¹ than at Summit. Therefore if the increase was due to firnification alone, one would not expect the two H_2O_2 records to be so nearly parallel when they are plotted on the same timescale (Fig. 3).

The high-resolution continuous measurements of the Summit

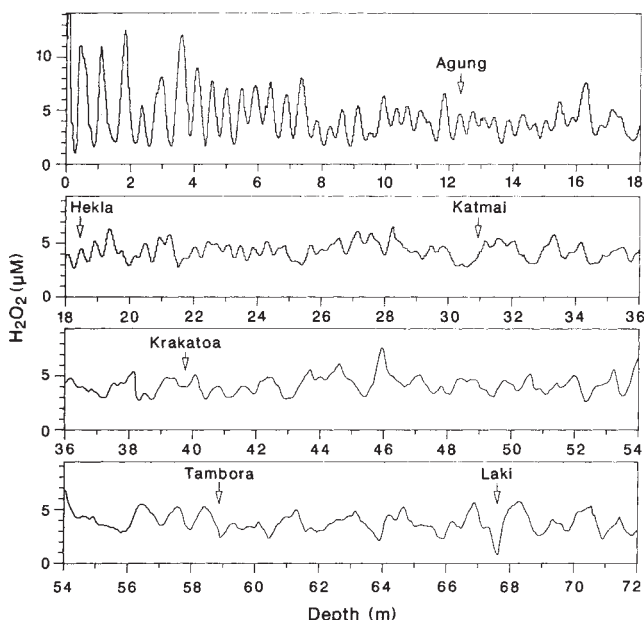


FIG. 1 Hydrogen peroxide concentrations in the firn part of the Summit core. The depths corresponding to dates of large volcanic eruptions are marked.

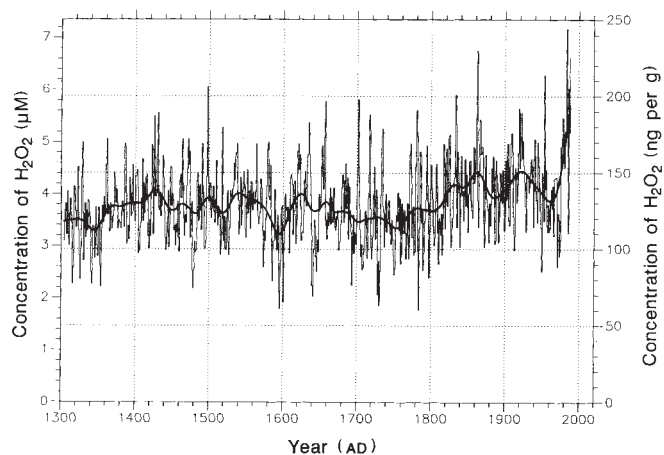


FIG. 2 Annual mean H_2O_2 concentrations during the past 700 years in the Summit ice core.

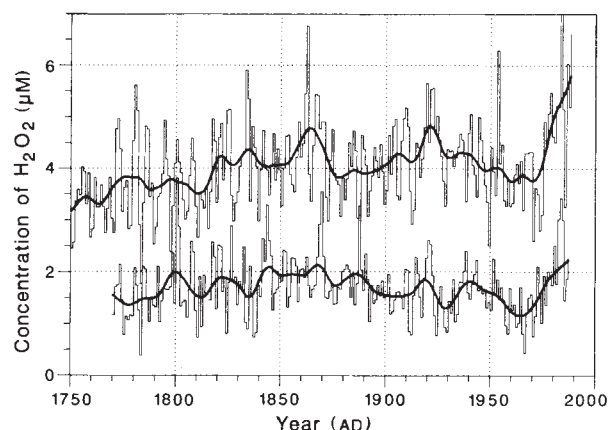


FIG. 3 Comparison of annual mean H_2O_2 concentrations from Summit (upper trace) and Dye 3 (lower trace).

core show that on a timescale of hundreds of years, chemical decomposition of H_2O_2 occurs in the ice. Below the pore close-off depth, thin layers of low H_2O_2 concentration start to develop and become more and more frequent with increasing depth. These 'negative peaks' have wavelengths of typically 10 cm. They must have been formed only after deposition in the ice, as wavelengths of 10 cm are smoothed out completely by firnification. A gaussian low-pass filter with diffusion length 12 cm attenuates signals of wavelength $\lambda = 10$ cm by a factor of 5×10^{-13} , seasonal cycles ($\lambda = 23$ cm) by 5×10^{-3} , and signals of $\lambda = 1$ m by only 0.75. These negative H_2O_2 peaks carry important information about the long-term stability of H_2O_2 in ice, but quantitatively they do not alter the annual mean concentrations of Fig. 2 by more than 5%. Low H_2O_2 concentrations are also associated with some of the volcanic events recorded in ice cores¹². Laj *et al.*¹² concluded that consumption of H_2O_2 by oxidation of volcanic SO_2 in Greenland precipitation is responsible for the decreased H_2O_2 levels. They based their conclusions on a limited data set of four core sections from a site near Summit, containing the volcanic deposits of the eruptions of Hekla (1947), Katmai (1912), Tambora (1815) and Laki (1783). Their findings are only partially confirmed by our continuous record. The large Icelandic Laki eruption is associated with the lowest H_2O_2 concentrations in the firn (Fig. 1). This ' H_2O_2 hole' is wide enough to be interpreted as a depletion of H_2O_2 in precipitation by oxidation of SO_2 , because it contains only wavelengths that survive firnification. But this is the only case where such an interpretation is unambiguous. The H_2O_2 depletions associated with other prominent volcanic events, such as Hekla, Katmai, Tambora, Krakatoa (1883) and Agung (1947), are no deeper than those found in other core sections as well. Laj *et al.*¹² deduced atmospheric SO_2 oxidation rates from the depletion of H_2O_2 and corresponding sulphate increases in their ice cores. We feel that this interpretation is going too far, as smoothing of H_2O_2 in the firn is neglected and the assumed H_2O_2 losses associated with volcanic events cannot be distinguished from other fluctuations of H_2O_2 in the ice.

An atmospheric increase of hydrogen peroxide in remote regions as a result of human activities is expected from model calculations. Thompson *et al.*¹³ predict a global increase of H_2O_2 of up to 33%, depending on the model scenario, between 1985 and 2035. The parameters governing the magnitude of the increase are the emission rates of NO_x , CH_4 and CO, the degree of depletion of stratospheric ozone and the degree of global warming affecting the water vapour content of the atmosphere. Nitrogen oxides, NO_x , tend to reduce the concentration of peroxide, whereas hydrocarbons, stratospheric ozone depletion and increasing water vapour enhance its production. Our data

indicate that human activities have probably already led to an increase in atmospheric hydrogen peroxide in Greenland. □

Received 1 October 1990; accepted 11 April 1991.

1. Logan, J. A., Prather, M. J., Wofsy, S. C. & McElroy, M. B. *J. geophys. Res.* **86**, 7210–7254 (1981).
2. Kleinman, L. I. *J. geophys. Res.* **91**, 10889–10904 (1986).
3. Perket, S. A., Jones, B. M., Brice, K. A. & Eggleston, A. E. *Atmos. Environ.* **13**, 123–137 (1979).
4. Möller, D. *Atmos. Environ.* **23**, 1625–1627 (1989).
5. Neftel, A., Jacob, P. & Klockow, D. *Nature* **311**, 43–45 (1984).
6. Lazrus, A. L., Kok, G. L., Gitlin, S. N., Lind, J. A. & McLaren, S. E. *Analyt. Chem.* **57**, 917–922 (1985).
7. Hwang, H. & Dasgupta, P. K. *Analytica chim. Acta* **170**, 347–352 (1985).
8. Dasgupta, P. K. & Hwang, H. *Analyt. Chem.* **57**, 1009–1012 (1985).
9. Sigg, A. & Neftel, A. *Ann. Glaciol.* **10**, 157–162 (1988).
10. Johnsen, S. J. *Int. Ass. Hydrol. Sci. Publ. No.* 118, 210–219 (Grenoble, 1977).
11. Herron, M. M. & Langway, C. C. *J. Glaciol.* **25**, 373–385 (1980).
12. Laj, P., Drummey, S. M., Spencer, M. J., Palais, J. M. & Sigurdsson, H. *Nature* **346**, 45–48 (1990).
13. Thompson, A. M., Owens, M. A. & Stewart, R. W. *Geophys. Res. Lett.* **16**, 53–56 (1989).

ACKNOWLEDGEMENTS. We thank H. Oeschger and B. Stauffer for the opportunity to do this work and for their encouragement, and H. Clausen for providing firn samples of the Crête Core. This work was supported by the University of Bern and the Swiss office for education and science through the COST 611 project 'Eurocore'.

Melt generation during rifting in the North Sea

Dave Latin* & Frances G. Waters†

* Bullard Laboratories, University of Cambridge, Madingley Road, Cambridge CB3 0EZ, UK

† Department of Earth Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EQ, UK

IN a forthcoming paper, McKenzie and O'Nions¹ invert rare-earth element concentrations in mid-ocean-ridge basalts to calculate the variation of melt fraction with depth in the asthenospheric mantle beneath ocean ridges. They find that melting persists to greater depths (about 80 km) than would be predicted by parameterizations of melting experiments² (about 50 km). This result has important implications for predicting the volume and composition of melt produced from asthenosphere of normal potential temperature during continental rifting. The North Sea rift is a good location to test models for melt generation because the physical conditions of rifting are well constrained^{3–5}. Using the distribution of melt fraction with depth derived in ref. 1, we calculate the volume and composition of melt produced from the asthenosphere in the most stretched part of the North Sea (the Forties region). When this melt is mixed in approximately equal proportions with an alkaline melt, similar to that produced from metasomatized lithosphere by very small amounts of stretching on the rift flanks, it accounts for the volume and the elemental and isotopic composition of the basalts in the most extended region.

Using parameterizations of batch melting experiments on dry peridotite, McKenzie and Bickle² have demonstrated that the average thickness and composition of the oceanic crust can be explained by adiabatic upwelling of asthenosphere with a potential temperature of 1,280–1,300 °C. Variations in thickness and composition of the oceanic crust are attributed to variations in the potential temperature of the convecting upper mantle^{2,9}. The experimental parameterizations predict that asthenosphere of normal potential temperature will not start to melt until it has risen to a pressure of less than 1.5 GPa (~50 km). The lithosphere consists of a rigid, conductive mechanical boundary layer (MBL), above a thermal boundary layer in which both conduction and advection control heat transport, and it overlies the purely advective asthenosphere². It is the thickness of the MBL before rifting that determines the pre-rift position of the geotherm relative to the solidus^{2,5,8,10}. The equilibrium lithosphere thickness of many extensional sedimentary basins is inferred to be similar to that in the oceans (~120–125 km)^{4,11}, implying an equilibrium MBL ~100 km thick. Thus at normal

TABLE 1 Observed and calculated melt compositions

	Forties basalt	Asthenospheric melt	Metasomatic melt	Flank melts	
				Nephelinite	Ultrapotassic
La	43.38	10.93	76.02	73.67	79.50
Ce	91.26	30.04	152.48	150.47	158.16
Pr	10.61	4.53	16.69	16.50	17.37
Nd	43.48	20.63	66.32	63.48	66.99
Sm	7.57	5.76	9.38	9.56	10.53
Eu	2.33	1.85	2.81	2.72	3.11
Gd	6.67	6.02	7.32	7.48	8.73
Dy	5.24	5.55	4.92	5.00	6.73
Ho	0.90	1.11	0.69	0.84	1.10
Er	2.41	2.86	1.96	2.18	2.87
Yb	1.88	1.88	1.88	1.88	2.14
Lu	0.27	0.26	0.29	0.23	0.31
SiO ₂	48.20	48.37	48.03	44.29	40.74
TiO ₂	2.53	1.78	3.28	3.41	4.24
Al ₂ O ₃	13.01	14.26	11.76	12.05	15.42
FeO	11.12	10.14	12.10	10.63	13.38
MgO	10.25	12.56	7.94	7.52	9.26
CaO	12.74	10.38	15.10	14.52	8.15
Na ₂ O	2.15	2.51	1.79	2.33	0.98
K ₂ O	1.01	0.24	1.78	0.87	4.57
¹⁴³ Nd/ ¹⁴⁴ Nd $\pm 2\sigma$ (at 150 Myr ago)	0.512655 $\pm$ 0.000022	0.512933	0.512646	0.512586 $\pm$ 0.000022	0.512678 $\pm$ 0.000026
Number in average	12			3	7

REE concentrations are in p.p.m.; oxide concentrations are in weight %.

potential temperatures the stretching factor β (defined as initial thickness/stretched thickness) must exceed 2.5 for melting to occur (see Fig. 1). The Jurassic-Cretaceous North Sea rift system, which seems to have formed by much smaller degrees of extension³⁻⁸ ($\beta \leq 2$), was nevertheless accompanied by magmatism.

The locus of magmatic activity in the North Sea was at the junction of the three main rift arms, where a thickness of ~ 1 km of alkali basalt was erupted over $\sim 15,000$ km² to produce the

Forties volcanic province^{5,8,12}. The volume of intrusive material in the Forties region is not known, but the relatively primitive nature of the basalts (MgO ~ 10 wt%) suggests it is unlikely that there are any large gabbroic plutons. This region of the North Sea seems to have been stretched by a factor of ~ 2 (refs 3, 4 and 7). Elsewhere, on the rift flanks and in sub-basins, small volumes of distinctive, highly alkaline, mafic (MgO > 8 wt%) nephelinite and ultrapotassic (K₂O ≥ 3 wt%) rocks, rich in primary hydrous phases (such as phlogopite and kaersutite), were extruded and intruded⁵. These off-axis magmas occur in areas where β is generally less than 1.2 and may be interpreted as melts produced on volatile-controlled solidi in metasomatized regions of the MBL⁵. The Forties basalts contain only minor amounts of primary volatile-bearing phases and are much more important in terms of volume than the off-axis magmas, therefore a component of the Forties basalts with a source in the asthenosphere seems likely.

If part or all of the melt erupted in the Forties province was produced by melting asthenosphere, and if the observations are to be reconciled with the experimental parameterizations, then either the estimates of extension are too low, or the potential temperature was higher than normal ($\sim 1,380$ °C), or the MBL was only ~ 70 km thick immediately before Jurassic rifting. There is little evidence to suggest that β exceeds 2 in the Forties region of the North Sea^{3,4,7}. Although a mantle plume may have been present during rifting, there is no evidence for regional uplift, on the scale required by White and McKenzie¹⁰, before rifting, and the 'doming' reported by Ziegler¹³ is a relatively localized and short-lived phenomenon⁷. Little is known about the position of the geotherm immediately before Jurassic extension, but a reduced MBL thickness (70 km) at that time requires there to have been considerable extension during the Triassic ($\beta > 2$); there is no evidence of this^{6,7,13}. In short, none of these three propositions seems particularly attractive.

Alternative possibilities are that either the solidus position used in the parameterizations² is not correct, or models based on batch melting experiments do not accurately reflect melting processes in the mantle, which are more likely to be 'dynamic'^{2,9,14}. The solidus position could be in error for two reasons. First the asthenosphere may not be completely free of volatiles¹⁵, which would allow small amounts of melt to be produced at lower temperatures or higher pressures than those

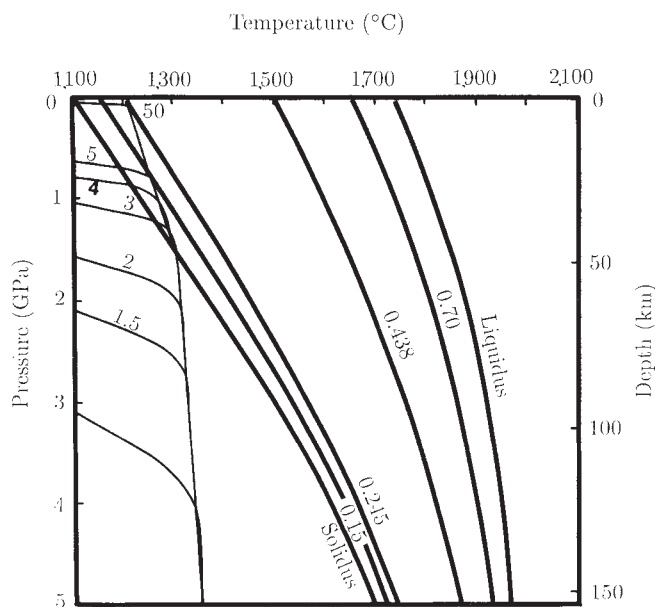


FIG. 1 Adiabatic upwelling caused by stretching a convective geotherm generated for a mechanical boundary layer initially 100 km thick (corresponding to a lithosphere ~ 125 km thick), overlying a convecting interior with a potential temperature of $1,280$ °C and a kinematic viscosity of 4×10^{15} m² s⁻¹. Stretched geotherms (thin lines) are labelled with values of β . Curves between the solidus and liquidus (thick lines) were generated from the parameterization of melting experiments² and are labelled with their melt fraction. Note that the solidus is not intersected until $\beta \approx 2.5$.

of the dry solidus. Second, the anhydrous melting experiments might be inaccurate. The 'dry' solidus, used in the theory, was determined by fitting a line between those experiments in which melt (glass) was observed and those in which it was not. It is well known that it is difficult to identify trace amounts of melt in an experimental charge, and the solidus may therefore reflect the onset of 'experimentally visible' melting rather than the true onset of melting. Although both arguments may be valid, we suspect that the largest effects arise from uncertainty about the details of the melting process. McKenzie and Bickle² took care to point out that a model based, out of necessity, on batch melting experiments might not accurately predict the generation of melt during Rayleigh melting. It is not clear that the relationship between the melt fraction (X) and T' (a dimensionless temperature² which describes the effective solidus overstep at a given P) will be the same for both batch and Rayleigh melting models. McKenzie and Bickle imply that for Rayleigh melting X may increase rather more slowly with T' than it does for batch melting. If this is the case, then a higher potential temperature is required to produce the average oceanic crust (melting must start at greater depths, and so the normal potential temperature of the asthenosphere must be higher than 1,280–1,300 °C). Until recently it seemed that these problems would remain unresolved.

In their forthcoming paper, McKenzie and O'Nions¹ use inverse theory to determine melt distributions from rare-earth element (REE) concentrations in relatively primitive basalts. Assuming a knowledge of the REE concentration of the melting region and its mineralogy, the partition coefficients for REEs between solid and melt and the REE concentrations of a suite of basaltic rocks, and using a Rayleigh melting model, McKenzie and O'Nions obtain a variation for X with depth below the top of the melting region. In this way the melt distribution with depth below 'normal' oceanic spreading centres (those at water depths ≥ 2 km) is obtained from the REE concentrations in an average mid-ocean-ridge basalt from these areas (N-MORB; see Fig. 2). The REE inversion predicts the correct amount of melt, but the variation in X with depth is different from that predicted by parameterizations for a potential temperature of 1,300 °C.

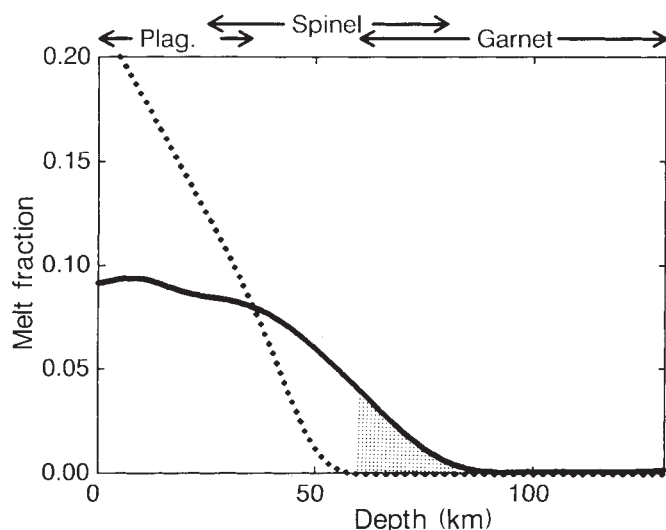


FIG. 2 The variation of melt fraction with depth in the asthenosphere upwelling beneath spreading ocean ridges. Solid line shows the melt distribution calculated by inversion of rare earth element distributions in N-MORB¹. The upper-mantle composition and mineralogy used in this calculation are given by McKenzie and O'Nions¹. Dotted line shows the melt distribution calculated from the parameterization² assuming adiabatic upwelling and a potential temperature of 1,300 °C. Shaded area indicates the region of asthenospheric melt generation (60–80 km) in the most stretched part of the North Sea. Also shown are the stability ranges for plagioclase, spinel and garnet peridotite.

Most noticeable is melt generation at depths of ~ 80 km, whereas the models predict that none should occur at depths > 50 km. The REE content of average N-MORB clearly requires some input from depths between 60 and 80 km. The difference between the two models is important if, as we suspect, the inversion scheme¹ provides a more accurate picture of the melt distribution resulting from dynamic (rather than batch) melting.

If we accept that the melt distribution obtained from the REE inversion accurately reflects the melting path taken by adiabatically upwelling asthenosphere of normal potential temperature, then we can use this result to predict the amount and REE composition of melt produced during stretching of the lithosphere. There will now be no difficulty in generating some melt from the asthenosphere if β is less than 2.5. Knowing the initial thickness of the lithosphere in the North Sea (~ 120 – 125 km), and the stretching factor in the Forties region ($\beta \sim 2$), then the asthenosphere must rise to a depth of ~ 60 km during extension. It is clear from Fig. 2 that melting will start as the asthenosphere rises to ~ 80 km and will reach a maximum melt fraction of just under 0.05 at 60 km. This melt distribution can then be used to calculate the thickness and composition of the asthenospheric melt produced.

The calculated thickness of melt is 0.66 km, which is slightly less than the observed thickness (~ 1 km). The predicted REE composition is compared with the average Forties basalt composition in Fig. 3. Although the fit is quite good for the middle and heavy REEs, the asthenospheric melt is considerably lower in light REEs. But if the asthenospheric melt mixes with some light-REE-enriched melts from metasomatized regions of the lithosphere, the fit can be improved. This melt may be similar in elemental and isotopic composition to the highly alkaline rocks produced at small β values on the rift flanks. Table 1 shows that a 'metasomatic melt' enriched in light REEs, calculated from the difference between the Forties basalts and the asthenospheric melt, compares well with the composition of the 'flank' melts. If the lithospheric contribution is similar to the off-axis nephelinites and ultrapotassic rocks then ~ 0.66 km of this metasomatic melt is required, giving a total melt thickness of 1.32 km, which is slightly greater than the observed thickness of ~ 1 km, although the latter is not tightly constrained. McKenzie and O'Nions¹ inverted the Forties data but were unable to separate asthenospheric and lithospheric melt contributions and so could not reproduce the observed compositions without assuming a light-REE-enriched amphibole-bearing mantle source. The approach taken here is different in that we calculate

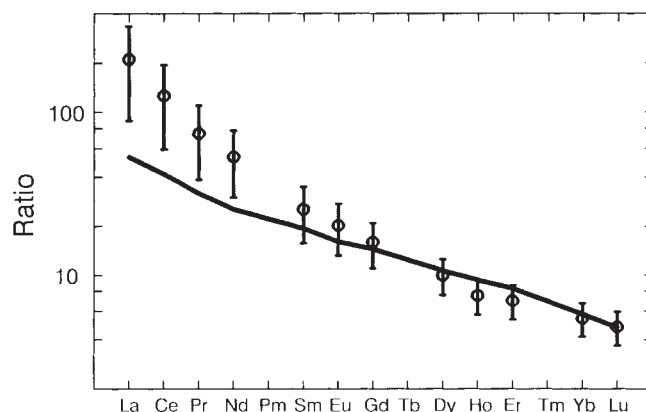


FIG. 3 Rare-earth element concentrations, calculated from the melt distribution below 60 km shown in Fig. 2 (solid line), compared with the observed REE concentrations in the Forties basalts (circles) with $\text{MgO} \geq 6$ wt% (see Table 1). Error bars show the variation within the average Forties dataset. Concentrations are normalized to a depleted mantle composition estimated by assuming that the upper 40% of the whole mantle was depleted during formation of the Earth's crust.¹

the asthenospheric melt (from the N-MORB melt distribution; no amphibole is required) and then improve the fit with the observed data by mixing in an enriched (lithospheric) melt.

Table 1 also shows the predicted average major element composition of the asthenospheric melt, calculated from the most recent version of the experimental parameterization¹⁶ using the melt fractions obtained from the inversion. The major element composition of the metasomatic melt has been calculated from the difference between the predicted and Forties basalt compositions. Comparison with the flank melts implies that the metasomatic melt may require some mixture of both types (nephelinitic and ultrapotassic). In this instance, it is difficult to be precise, because all of the observed major element compositions have been affected by fractional crystallization.

Finally we consider the isotopic composition of the observed and predicted melts. The least-altered Forties basalt erupted ~150 Myr ago with a ¹⁴³Nd/¹⁴⁴Nd ratio of 0.512655. Assuming that the asthenospheric melt had a depleted Nd isotopic composition at that time (calculated assuming that depleted mantle has ¹⁴³Nd/¹⁴⁴Nd = 0.513151 and Sm/Nd = 0.3669 today¹⁷), we can calculate the Nd isotopic composition of the metasomatic melt (see Table 1). Note that because equal mixing proportions have been assumed, and the Nd content of the metasomatic melt is much higher than that of the asthenospheric end member, the calculated ¹⁴³Nd/¹⁴⁴Nd ratio of the metasomatic melt (0.512566) is similar to the ¹⁴³Nd/¹⁴⁴Nd ratio of both the Forties basalts and the off-axis melts. We have not presented the Pb and Sr isotope data because the partition coefficients of Pb and Sr, as well as their upper-mantle concentrations, are relatively poorly known, and so these elements are not yet used in the inversion¹.

The model presented above is not unique, depending as it does on a knowledge of the asthenospheric composition, mineralogy and REE partition coefficients, as well as the assumption that a lithospheric component is involved. Nevertheless, evidence for magmas produced by mixing of depleted asthenospheric and enriched lithospheric components during continental extension has often been discussed^{18–25}. Our model is consistent with these ideas, and provides a simple quantitative means of estimating the relative contributions of asthenosphere and lithosphere to rift-related magmatism in the North Sea. □

Received 20 March; accepted 8 May 1991.

- McKenzie, D. P. & O'Nions, R. K. *J. Petrol.* (in the press).
- McKenzie, D. & Bickle, M. J. *J. Petrol.* **29**, 625–679 (1988).
- Slater, J. G. & Christie, P. A. F. *J. geophys. Res.* **85**, 3711–37 (1980).
- Barton, P. & Wood, R. *Geophys. J. R. astr. Soc.* **79**, 987–1022 (1984).
- Latin, D. M., Dixon, J. E., Fitton, J. G. & White, N. in *Tectonic Events Responsible for Britain's Oil and Gas Reserves* (eds Hardman, R. F. P. & Brooks, J.) *Geol. Soc. spec. Publ.* **55**, 207–227 (1990).
- White, N. in *Tectonic Evolution of the North Sea Rifts* (eds Blundell, D. & Gibbs, A.) 217–239 (Oxford Science Publications, 1990).
- White, N. & Latin, D. *Tectonophysics* (in the press).
- Latin, D. & White, N. *Geology* **18**, 327–331 (1990).
- Klein, E. M. & Langmuir, C. H. *J. geophys. Res.* **92**, 8089–115 (1987).
- White, R. S. & McKenzie, D. P. *J. geophys. Res.* **94**, 7685–729 (1989).
- Slater, J. G., Parsons, B. & Jaupart, C. *J. geophys. Res.* **96**, 11535–52 (1991).
- Woodhall, D. & Knox, R. W. *Bull. geol. Survey* **70**, 34–56 (1979).
- Ziegler, P. A. in *Tectonic Evolution of the North Sea Rifts* (eds Blundell, D. & Gibbs, A.) 1–36 (Oxford Science Publications, 1990).
- Elliott, T. R., Hawkesworth, C. J. & Grönvald, K. *Nature* **351**, 201–206 (1991).
- Michael, P. J. & Chase, R. L. *Contrib. Min. Petrol.* **96**, 245–63 (1987).
- Watson, S. & McKenzie, D. *J. Petrol.* **32**, 501–537 (1991).
- Miller, R. G. & O'Nions, R. K. *Nature* **314**, 325–330 (1985).
- Cox, K. G. in *Continental Basalts and Mantle Xenoliths* (eds Hawkesworth, C. J. & Norry, M. J.) 139–158 (Shiva, Nantwich, Cheshire, 1983).
- Menzies, M. A. & Hawkesworth, C. J. (eds) *Mantle Metasomatism* (Academic, London, 1987).
- McKenzie, D. *Earth planet. Sci. Lett.* **95**, 53–72 (1989).
- Hawkesworth, C. J., Kempton, P. D., Rogers, N. W., Ellam, R. W. & van Calsteren, P. W. *Earth planet. Sci. Lett.* **96**, 256–268 (1990).
- Thompson, R. N. *et al.* *Earth planet. Sci. Lett.* **98**, 139–153 (1990).
- Waters, F. G., Cohen, A. S., O'Nions, R. K. & O'Hara, M. J. *Earth planet. Sci. Lett.* **97**, 241–255 (1990).
- Ellam, R. M. & Cox, K. G. *Earth planet. Sci. Lett.* (in the press).
- Wilson, B. M. & Downes, H. *J. Petrol.* **32**, 811–849 (1991).

ACKNOWLEDGEMENTS. We thank D. McKenzie and K. O'Nions for access to their inversion program and for useful discussion, M. Bickle, K. Cox, J. Dixon, S. Watson, N. White and R. White for comments, and R. Ellam and T. Elliott for reviews. D.L. is supported by a NERC research fellowship.

A molecular genetic analysis of kinship and cooperation in African lions

C. Packer*, D. A. Gilbert†, A. E. Pusey* & S. J. O'Brien‡

* Department of Ecology, Evolution and Behavior, University of Minnesota, Minneapolis, Minnesota 55455, USA

† BioMolecular Research Department, Washington Research Center, W. R. Grace & Co.-Conn., Columbia, USA

‡ Laboratory of Viral Carcinogenesis, National Cancer Institute, Frederick Cancer Research and Development Center, Frederick, Maryland 21702-1201, USA

AFRICAN lions live in complex social groups and show extensive cooperative behaviour^{1–10}. Here we describe a new application of DNA fingerprinting that unequivocally demonstrates the kinship structure of lion 'prides': female companions are always closely related, male companions are either closely related or unrelated, and mating partners are usually unrelated. The variability in relatedness among male coalition partners provides an important opportunity to test for the effects of kinship on cooperative behaviour¹¹. Paternity analysis reveals that male reproductive success becomes increasingly skewed as coalition size increases, and the tendency to form coalitions with non-relatives drops sharply with increasing coalition size. Thus males only act as non-reproductive 'helpers' in coalitions composed of close relatives.

Lion prides typically contain 2–9 adult females (range, 1–18), their dependent young and a coalition of 2–6 adult males (range, 1–9) that has entered the pride from elsewhere⁸. Incoming males kill or evict the dependent young of the prior coalition^{2,7,12,13}. Consequently, females resume sexual receptivity within days, show regular oestrus cycles and mate exclusively with the males of the new coalition by the time they conceive^{2,14}. Births are often synchronous within a pride² and cubs born less than one year apart comprise a 'cohort'. The resident males typically father only one cohort per pride^{4,7}. Two-thirds of all female cohorts are recruited into their mothers' prides, the remainder emigrating together to establish a new pride nearby⁷. By contrast, almost all males leave their natal pride before the age of 4 years and undergo a nomadic phase before gaining residence in a new pride⁷. Large male cohorts usually enter new prides intact, but cohorts of only one or two males often team up with single males from other prides before gaining residence^{3–8}.

Lions in the Serengeti National Park and Ngorongoro Crater, Tanzania have been studied continuously since the 1960s^{1,8,15}. The Serengeti ecosystem covers 25,000 km², the total lion population exceeds 3,000 and males disperse over the entire region⁷. More than two-thirds of the 118 males that have bred in our 2,000-km² study area originated in other parts of the Serengeti^{7,15}. By contrast, the 250-km² floor of the crater is a small isolated island of savannah. The current lion population is descended from 15 founders that survived an epizootic in 1962 and includes only 25–45 breeding adults¹⁵. Compared with the Serengeti population, the crater lions have significantly less allozyme heterozygosity, less restriction-fragment length polymorphism in major histocompatibility complex class I genes, and more sperm abnormalities^{15–18}.

Hypervariable DNA sequences were identified in the domestic cat and corresponding DNA probes were isolated¹⁹. In lions, these probes recognized multiple genetic loci with abundant DNA variation and were used in a DNA fingerprinting analysis^{20,21} of 193 animals from the Serengeti and 23 from the crater¹⁹. Data on matrilineal kinship come from long-term records⁸; parentage was also determined for 78 cubs in the Serengeti by DNA band matching. DNA analysis confirmed behavioural estimates of maternity for 77 of 78 cubs; the exception belonged

to a female primate of the assumed mother¹⁹. All 78 cubs were fathered by a member of the resident coalition¹⁹.

The extent of minisatellite band sharing (see legend to Fig. 1) was calibrated against known kinship to provide estimates of relatedness between individuals of uncertain parentage¹⁹. In the Serengeti population, band sharing is highly correlated with

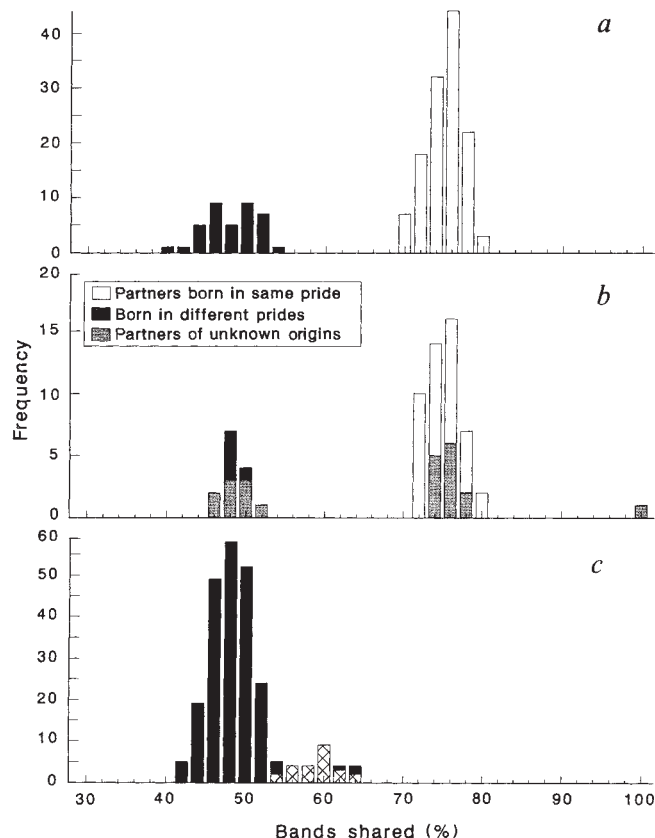


FIG. 1 Frequency distributions of minisatellite band sharing within or between different classes of individuals in the Serengeti. DNA is digested with restriction enzymes and fragments are separated according to relative molecular mass (M_r) by agarose gel electrophoresis. DNA fragments carrying different numbers of the hypervariable minisatellite repeat are seen as bands corresponding to varying M_r after hybridization with a radioactive minisatellite DNA probe. Percentage similarity or 'band sharing' between two individuals is $2F_{ab}/(F_a + F_b) \times 100\%$, where F_{ab} is the number of DNA fragments showing similar M_r and intensity carried by both individuals, F_a is the total number of fragments resolved in individual a and F_b the number resolved in b. All gels were scored by D.A.G., who did not know the genealogical relations of most individuals¹⁹. For overlapping distributions, the number of pairs showing a particular degree of band sharing is given by the height of the respective hatched region. *a*, Band sharing between female primate-mates (□) compared with band sharing of unrelated individuals born in different parts of the park (■). Data on female primate-mates are based on 63 females from 15 prides and include all pairwise combinations of females within each pride (for example, a pride of four females contributes six combinations of females). But each unrelated individual is included in only one combination (76 different individuals in 38 pairs). *b*, Band sharing between male coalition partners. Data are based on 45 males in 16 coalitions and include all pairwise combinations within each coalition. Data are plotted separately for partners known to have been born in the same pride (□), those known to have been born in different prides (■) and those that had entered the study area from elsewhere (▨). One pair of males of unknown origins shared 100% of their bands and are hence presumed to be identical twins. *c*, Band sharing between resident males and pride females. Data based on 44 males in 18 coalitions and 52 females in 15 prides, including all male-female combinations within each pride. Most males had no known kinship links to the females (■), but one coalition resided in a neighbouring pride that had split off from the males' natal pride 10 years before the males' births (▨).

relatedness but the relationship is not linear: individuals related by a coefficient of relatedness r of 0.5 share 79.6% (s.d. 2.2%) of their bands, animals related by $r = 0.25$ share 74.8% $\pm$ 2.2%, whereas those related by $r = 0.02$ –0.06 share only 62.0% $\pm$ 4.2%. Individuals assumed to be unrelated because they were born in prides with no known kinship links share 49.0% $\pm$ 3.3%. The bandsharing data thus distinguish three classes of relationship: kin related by $r \geq 0.125$, kin related by $r = 0.02$ –0.06 and non-relatives¹⁹. Band sharing declines more linearly with decreasing relatedness in the crater population and there is a greater variance in band sharing for a given degree of kinship. Crater lions related by $r = 0.5$ share 71.1% $\pm$ 11.1% of their bands, and those related by $r = 0.25$ share 59.1% $\pm$ 8.4% (note that these estimates of r do not consider the history of inbreeding in this population) and crater lions share 36.9% $\pm$ 4.2% with unrelated lions from the ancestral Serengeti population^{15,19}.

The minisatellite bandsharing data clearly reveal kinship within and between lion prides (Figs 1–3). In the Serengeti population, female primate-mates show far greater band sharing than do unrelated individuals (Fig. 1*a*). Band sharing between male coalition partners from the same natal pride is similar to female primate-mates, whereas partners born in different prides show band sharing typical of unrelated individuals (Fig. 1*b*); band sharing between partners of unknown origins clearly belongs to one distribution or the other (Fig. 1*b*). Resident males in the Serengeti prides are almost always unrelated to the pride females (Fig. 1*c*). When a cohort of daughters first splits from

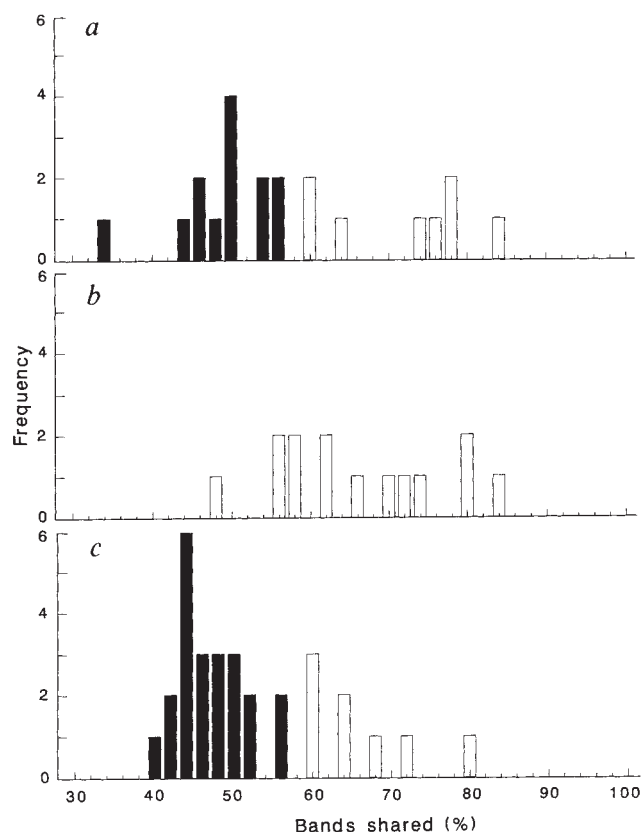


FIG. 2 Frequency distributions of band sharing in Ngorongoro Crater, plotted as in Fig. 1. *a*, Band sharing between female primate-mates (eight females in three prides; □) compared with band sharing of 22 individuals from crater prides with the fewest kinship links (■). *b*, Band sharing between male coalition partners (fourteen males in five coalitions). Partners in each coalition were born in the same pride. *c*, Band sharing between resident males and pride females (eleven males in four coalitions and six females in two prides). All of these males became resident outside their natal pride (■), but members of one coalition later returned to reside in their natal pride (□).

the parental pride, band sharing between the two prides is as close as within each pride (Fig. 3a). But band sharing between prides declines with decreasing kinship as the mothers die (maximum female lifespan is 17 years⁸) and the respective prides produce further cohorts of offspring fathered by different male coalitions (Fig. 3a). Even though the crater population is small and isolated, crater females share more bands with pridemates than with individuals from the most distantly related prides (Fig. 2a); male coalition partners born in the same pride share a proportion of bands similar to female pridemates (Fig. 2b); males resident in non-natal prides share fewer bands with the females than do males resident in their natal pride (Fig. 2c); and band sharing declines with time after prides split (Fig. 3b).

Males cooperate by patrolling the pride's territory and chasing out all male intruders. Because all coalition partners participate in group defence against other coalitions, larger coalitions are better able to gain residence in a pride, maintain residence for longer periods, and gain access to more females over their lifespans^{4,8}. A resident coalition is highly effective in preventing extra-pride males from fathering cubs: the paternity analysis demonstrates that the resident males father all cubs conceived during their tenure¹⁹. Long-term data show that larger coalitions father more surviving offspring per capita than small coalitions^{4,8}. An increase in coalition size of one additional male results in an additional 0.64 surviving cubs per male⁸.

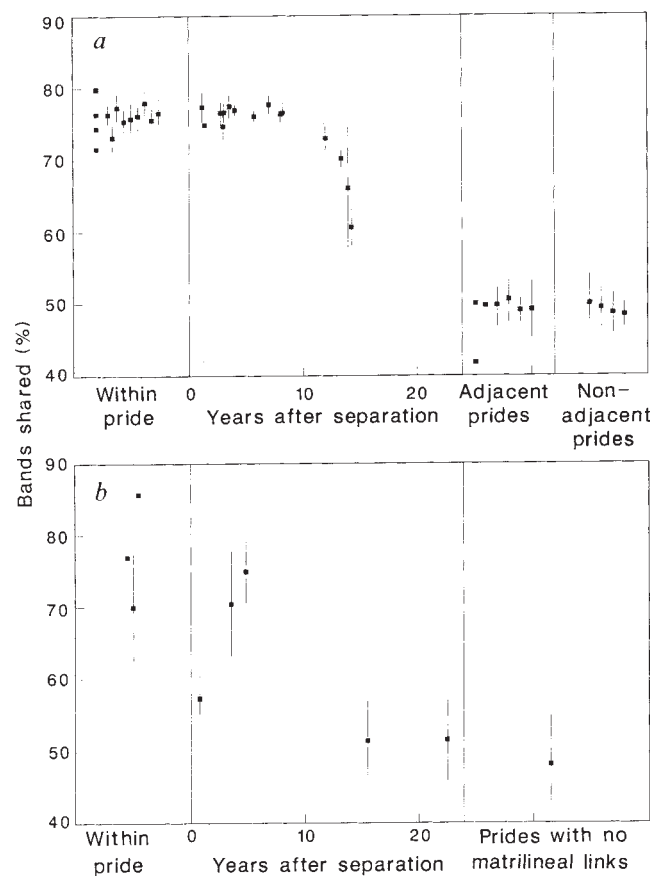


FIG. 3 Degree of band sharing within and between prides. Each point represents comparisons within a pride or between two prides. Where comparisons are made between prides of common female ancestry, the data are plotted against the number of years that have elapsed after the prides separated. The mean and s.d. are plotted across all pairwise combinations for each comparison. *a*, Serengeti. Data on adjacent and non-adjacent prides only include prides with no known kinship links (either through males or females) since the beginning of the study. All comparisons are of females. *b*, Ngorongoro Crater. Comparisons between crater prides unavoidably involve numerous kinship links: most contemporary prides are descended from a founding group of four females and there has been frequent exchange of males between all five prides¹⁵. Between-pride comparisons include both sexes.

Although a significant proportion of males form coalitions with unrelated companions (Fig. 1b), they do so only under specific circumstances. As coalition size increases, the proportion of coalitions containing non-relatives drops sharply (Fig. 4a). Single males only group to form pairs and trios; all larger coalitions are composed entirely of close relatives. Because per capita reproductive success increases with increasing coalition size, why is it that unrelated males do not continue to join to form larger coalitions? The paternity analysis reveals that it is probably because within-coalition variance in individual reproductive success increases markedly with increasing coalition size (Fig. 4b). Partners had similar reproductive success in all coalitions of only two males, but at least one male failed to breed in each of the larger coalitions.

Cooperation between non-relatives is most likely to evolve when all group members can breed²²⁻²⁴. Although each member

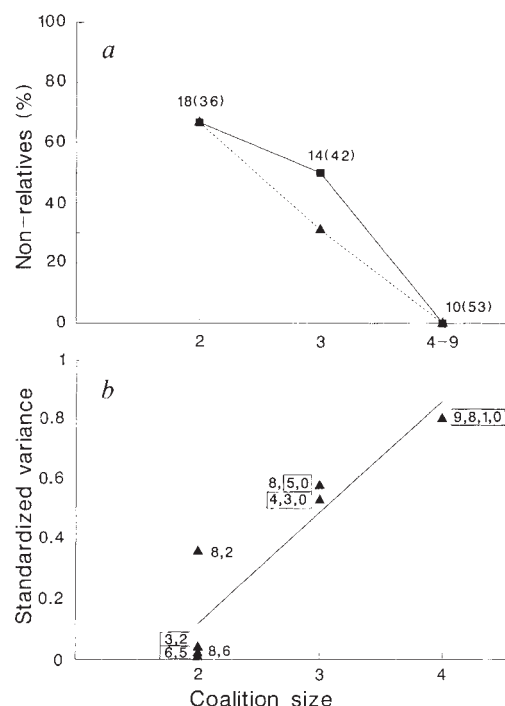


FIG. 4 *a*, Percentage of coalitions containing non-relatives (■) and of males lacking related partners (▲). Data include all coalitions for which we have demographic or bandsharing information. Coalitions containing unrelated males are significantly smaller than those composed exclusively of relatives ($z=3.122$, $n_1=19$, $n_2=23$, $P<0.002$, two-tailed, Mann-Whitney U -test; if the analysis is restricted to coalitions for which we have demographic data, $z=3.300$, $n_1=14$, $n_2=18$, $P<0.001$). Numbers by each point give the number of coalitions of that size; numbers in parentheses give the number of males. The percentage of males lacking related partners declines even more sharply with increasing coalition size because four of the seven trios containing non-relatives were composed of two relatives and an unrelated third. In these cases, only one of the three males lacked a related partner. *b*, Standardized variance in individual reproductive success within each coalition. Numbers by each point indicate the number of cubs fathered by each male in that coalition and a box around these points indicates that the males are close relatives. For example, all four males in the coalition of four were relatives; the most successful male in this coalition fathered nine cubs, and the remaining males fathered eight, one and zero cubs respectively. For the regression of standardized variance against coalition size, $r^2=0.8370$, $P<0.001$. Probability for the regression is found by simulation. Simulations were performed by random generation of reproductive success for the 18 males according to the observed distribution of coalition sizes and the observed number of cubs fathered by each of the seven coalitions. The observed regression of standardized variance against coalition size was exceeded in only 2 of 10,000 runs.

of a pair can breed, one or more members of each larger coalition may best be viewed as non-breeding helpers. Their presence increases the reproductive success of their companions (through augmenting coalition size), but they have little personal reproductive success. Such behaviour could only enhance their genetic representation in subsequent generations if it increased the reproductive success of close relatives¹¹. A relatively low proportion of males in trios lack any kinship connections with their companions (Fig. 4a). Many trios containing non-relatives consist of a related pair with an unrelated third. The paternity data include one such trio and the non-breeder in this group was closely related to one of the breeders (Fig. 4b). His failure to breed was ameliorated by his relative's success.

Within-coalition variance in male reproductive success results from two components. First, males typically father entire litters. A male will attempt to monopolize a female throughout her 2–4-day oestrus by remaining nearby and preventing his coalition partners from mating with her^{5,14}. As in other species²⁵, this behaviour ensures that the consorting male fathers her entire litter: mixed paternity occurred in only one of 24 litters containing two or more cubs¹⁹. Second, within-coalition variance in the number of litters fathered by each male increases significantly with increasing coalition size ($P = 0.0084$, $n = 44$ litters; probability determined as described in legend to Fig. 4b). Breeding within each pride is typically synchronous so that several females are often in oestrus at once^{2,26}. But larger coalitions do not preferentially reside in larger prides⁸, thus the number of males in a large coalition more often exceeds the number of available oestrous females.

Solitary males spend long periods selecting potential companions⁷ and behavioural studies indicate that mating success is more disparate within coalitions where partners are less closely matched in age or vigour⁵. The risk of becoming a non-breeder is greater in a trio, and males that form unrelated trios seem to be more sensitive to potential disparities in mating success than males that form pairs: the age difference between unrelated partners is significantly lower in trios (0.75 years) than in pairs (2.5 years, $P < 0.05$).

Previous explanations for cooperation among male lions did not require kinship to maintain the behaviour. Each male was assumed erroneously to father an equal proportion of offspring in all coalition sizes, and therefore to gain equally as per capita reproductive success increased with coalition size^{4,5,8}. But it is now clear that this is only true for small coalitions, kinship is essential for the maintenance of larger coalitions where reproduction is highly skewed.

In the Serengeti population, the DNA band-sharing technique allowed us to measure whether companions, neighbours or mating partners were even distantly related. Although these data suggest that similar analyses could be conducted in populations that had not previously been studied in such detail, the relationship between relatedness and band sharing differed between the Serengeti and the nearby Ngorongoro Crater. Thus the degree of band sharing should be calibrated against an independent measure of kinship^{19,27}. □

Received 19 November 1990; accepted 3 April 1991.

- Schaller, G. B. *The Serengeti Lion* (University of Chicago Press, Chicago 1972).
- Bertram, B. C. R. *J. Zool.* **177**, 463–482 (1975).
- Bertram, B. C. R. in *Growing Points in Ethology* (eds Bateson, P. P. G. & Hinde, R. A.) 281–301 (Cambridge University Press, Cambridge, 1976).
- Bygott, J. D., Bertram, B. C. R. & Hanby, J. P. *Nature* **282**, 839–841 (1979).
- Packer, C. & Pusey, A. E. *Nature* **296**, 740–742 (1982).
- Packer, C. in *Ecological Aspects of Social Evolution* (eds Rubenstein, D. I. & Wrangham, R. W.) 429–451 (Princeton University Press, Princeton, 1986).
- Pusey, A. E. & Packer, C. *Behaviour* **101**, 275–310 (1987).
- Packer, C. et al. in *Reproductive Success* (ed. Clutton-Brock, T. H.) 363–383 (University of Chicago Press, Chicago, 1988).
- Packer, C., Scheel, D. & Pusey, A. E. *Am. Nat.* **136**, 1–19 (1990).
- Scheel, D. & Packer, C. *Anim. Behav.* **41**, 697–710 (1991).
- Hamilton, W. D. *J. theor. Biol.* **7**, 1–52 (1964).
- Packer, C. & Pusey, A. E. in *Infanticide: Comparative and Evolutionary Perspectives* (eds Hausfater, G. & Hrdy, S. B.) 31–42 (Aldine, New York, 1984).

- Hanby, J. P. & Bygott, J. D. *Anim. Behav.* **35**, 161–169 (1987).
- Packer, C. & Pusey, A. E. *Am. Nat.* **121**, 716–728 (1983).
- Packer, C. et al. *Conservation Biology* (in the press).
- O'Brien, S. J. et al. *Nat. geogr. Res.* **3**, 114–124 (1987).
- Wildt, D. E. et al. *Nature* **329**, 328–331 (1987).
- Yuhki, N. & O'Brien, S. J. *Proc. natn. Acad. Sci. U.S.A.* **87**, 836–840 (1990).
- Gilbert, D., Packer, C., Pusey, A. E. & O'Brien, S. J. (submitted).
- Jeffreys, A. J., Wilson, V. & Thein, S. L. *Nature* **316**, 76–79 (1985).
- Burke, T. *Trends Ecol. Evol.* **4**, 139–144 (1989).
- Burke, T., Davies, N. B., Bruford, M. W. & Hatchwell, B. J. *Nature* **338**, 249–251 (1989).
- Strassman, J. E. et al. *Nature* **342**, 268–270 (1989).
- Rissing, S. W., Pollock, G. B., Higgins, M. R., Hagen, R. H. & Smith, D. R. *Nature* **338**, 420–422 (1989).
- Sherman, P. W. *Nature* **338**, 418–420 (1989).
- Packer, C. & Pusey, A. E. *Anim. Behav.* **31**, 334–340 (1983).
- Lynch, M. *Molec. Biol. Evol.* **5**, 584–599 (1988).

ACKNOWLEDGEMENTS. We thank M. Bush, J. Grinnell, L. Herbst, M. Jago, D. Janssen, C. Malima, S. Monfort and D. Wildt for assistance in collecting blood samples, and J. Avise and W. D. Hamilton for comments and discussion. NOAA center provided facilities for the DNA analysis and J. Martenson provided invaluable assistance. We thank the Tanzanian Government for permission and facilities. The field research was supported by the NSF, the University of Minnesota and the National Geographic Society.

Mitochondrial DNA analysis implying extensive hybridization of the endangered red wolf *Canis rufus*

R. K. Wayne* & S. M. Jenks†‡

* Department of Biology, University of California at Los Angeles, Los Angeles, California 90024, USA

† Department of Physiology, University of California at San Francisco, San Francisco, California 94143, USA

THE red wolf, previously endemic to the southeastern United States, declined precipitously in numbers after 1900 because of habitat destruction, predator control programmes, and hybridization with coyotes^{1,2}. Hybridization with coyotes probably occurred as these animals, which adjust well to agriculture, became numerous and moved eastwards^{1–4}. By 1970, red wolves existed only in extreme southeastern Texas and southwestern Louisiana (Fig. 1)². In 1967, red wolves were classified as endangered and a captive breeding programme was begun in 1974 after passage of the Endangered Species Act, about a year before they became extinct in the wild. Protein electrophoresis and morphometrics have been used to try to discriminate red wolves from hybrids and coyotes^{1,4,5}. But because the average substitution rate of mitochondrial DNA in mammals is much greater than that of nuclear genes⁶, mtDNA analysis is a more useful way of distinguishing closely related species. We have now analysed mtDNA restriction-enzyme sites and cytochrome *b* gene sequence variation in captive red wolves and in 77 canids sampled during the capture period. We also used the polymerase chain reaction to amplify and then sequenced mtDNA from red wolf skins collected before substantial hybridization of red wolves with coyotes is thought to have occurred. Phylogenetic analysis indicates that red wolves have either a grey wolf or coyote mtDNA genotype, demonstrating hybridization among these species. Thus, the red wolf is entirely a hybrid form or a distinct taxon that hybridized with coyotes and grey wolves over much of its previous geographical range. Our findings, however, do not argue against the continued protection of the red wolf.

We analysed red wolf mtDNA from three sources: the captive colony, blood samples collected between 1974 and 1976 in Texas and Louisiana, and skins collected between 1905 and 1930 (Fig. 1).

The four matrilineal lines represented in the captive red wolf population displayed a single mtDNA genotype. This genotype was

‡ Present address: Department of Psychology, University of California at Berkeley, Berkeley, California 94720 USA

TABLE 1 Genotype frequencies

	No. of restriction enzymes used				Total
	1	2	3	4	
Coyote-type	22	28	15	—	65 (84%)
Gray wolf-type (Northern)	—	1	—	4	5 (7%)
Gray wolf-type (Mexican)	4	1	—	2	7 (9%)
Total	26	30	15	6	77

Frequency of coyote-type and grey wolf-type genotypes in Texas and Louisiana canids from 1974 to 1976 (Fig. 1). Entries indicate the number of individuals classified as having a coyote or grey wolf genotype on the basis of restriction fragment patterns for one to four diagnostic restriction enzymes. Methods. DNA preparation and analysis follows our protocol described in refs 7 and 9. DNA samples were surveyed for restriction site variation using the following 4 restriction endonucleases: *BclI*, *BglII*, *EcoRV* and *HincII*. These enzymes produce diagnostic restriction fragment patterns unique to grey wolves (grey-wolf type) or coyotes (coyote-type). *HincII* digestion produces a restriction fragment pattern unique to captive red wolves and two wild-caught coyotes from Louisiana⁸. *BglII* digestion produces a restriction fragment pattern found only in Mexican grey wolves.

compared with those found in a sample of 327 coyotes and 276 grey wolves from throughout North America^{7,8}. To do this we used 21 restriction enzymes which revealed 138 independent restriction sites representing 670 nucleotides or roughly 4.0% of the canid mtDNA genome⁹. A phylogenetic tree relating a subset of these genotypes by maximum parsimony suggests that the genotype of the captive red wolf is grouped in a monophyletic clade including only coyote genotypes, and is indistinguishable from coyote genotype 32 found in two recent coyotes from Louisiana (Fig. 2a). Thus, the captive red wolves have a mtDNA genotype phylogenetically grouped with those of living coyotes.

Owing to sampling error or stochastic lineage extinction, however, the captive red wolf genotype may not be representative of those in the source population. Thus, we examined mtDNA from blood samples of 77 wild canids captured during the initial selection period (Fig. 1). These canids were classified on morphologic criteria as coyotes (58%), red wolf/coyote hybrids (31%), or red wolves (11%). Because the mtDNA yield in these samples was small we used only one to four diagnostic restriction enzymes^{7,8} and found that 84% of the samples had a coyote genotype, 7% had a northern grey wolf genotype, and 9% had a polymorphism found only in Mexican grey wolves (Table 1, Fig. 1). We were able to type 30 of the 77 canids with the enzyme *HincII*, and found that 23 (77%) had the restriction-fragment pattern that is diagnostic of captive red wolves (Table 1). Thus, the coyote-like genotype of captive red wolves was common in wild canids from the source population.

Morphologic and mtDNA classifications of the 1974–76 canids often do not correspond. For example, of the 12 canids with a grey wolf mtDNA genotype, one was identified as a red wolf, four as coyotes, and six as red wolf/coyote hybrids, and one as unknown. The high frequency of morphologic hybrids, the poor correspondence between mtDNA and morphologic classifications, and the presence of only coyote and grey wolf genotypes, suggest hybridization occurred between these two species in the source population before 1974. A less advanced hybrid zone between grey wolves and coyotes now exists in the Great Lakes region⁷. Our data also indicate that northern and Mexican grey wolf genotypes persisted in Texas canids a long time after the grey wolf had become extinct in the southeastern United States^{1,3}.

The apparent hybridization among Texas and Louisiana canids compelled us to examine samples from museum pelts of red wolves collected from 1905 to 1930, before extensive hybridization with coyotes is thought to have occurred (Fig. 1)¹. We amplified 398 base pairs of the cytochrome *b* gene from six pelts from five southeastern states using universal primers and

the polymerase chain reaction (Fig. 1)¹⁰. Sequences were compared with those obtained by the same method from recent blood samples of five coyotes, three grey wolves, a captive red wolf, a dog and a golden jackal (Table 2). The results confirm the conclusions suggested by restriction-site data⁷. Grey wolves and coyotes are very distinct, differing by 21–28 nucleotides or roughly 5.3–7.0% of their cytochrome *b* sequence. Within each species, no more than seven substitutions separate genotypes. The six historical red wolf sequences are similar or identical to those of the grey wolf or coyote. A phylogenetic analysis of the cytochrome *b* sequence data shows two significant monophyletic clades found in all 1,000 bootstrap replicates of our data¹¹: one clade containing grey wolf and three red wolf genotypes and the other containing coyote and two red wolf genotypes (Fig. 2b). Thus, red wolf genotypes are classified with those of either the coyote or grey wolf.

Two situations could account for our results. First, the red wolf could have been a distinct species with unique mtDNA genotypes that were missed in our survey or had become extinct through genetic drift. This species then hybridized with both coyotes and grey wolves over much of its geographical range. Second, the red wolf phenotype could have derived entirely from hybridization between coyotes and grey wolves¹². Red wolves are intermediate in size and may be difficult to distinguish morphologically from the other two species^{1,2,4}. The red wolf could have been a southeastern subspecies of the grey wolf that was morphologically but not genetically (by our criteria) distinct from other grey wolves. This subspecies then hybridized with coyotes that were numerically increasing throughout much of the red wolf's geographical range. Arguing against red wolves' constituting a distinct subspecies, are the probable high rates of gene flow from adjacent populations of grey wolves living in similar habitats that would tend to obliterate the distinctive morphology of the red wolf subspecies. In fact, gene flow has apparently stifled genetic differentiation among even very distant

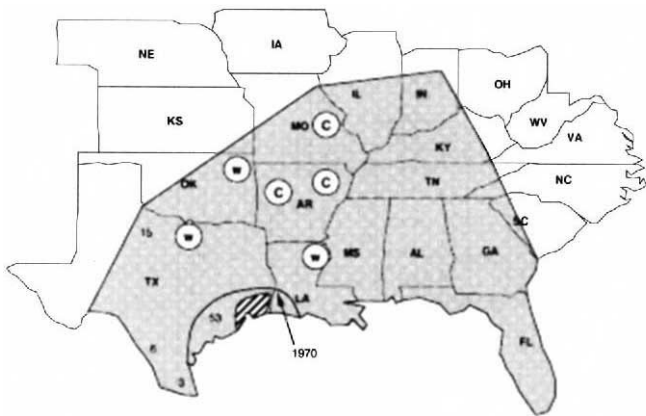


FIG. 1 Geographical range and sampling localities for red wolves. The shaded area indicates the distribution of red wolves circa 1700 (ref. 16) and the smaller bounded area in southeastern Texas the geographical range circa 1970, near the initiation time of the red wolf recovery programme. Map numbers indicate the location and number of Texas and Louisiana canids from 1974 to 1976 analysed; circled letters indicate historical samples and their genotype (C, coyote; W, grey wolf). The striped area in extreme southeastern Texas includes locations of 12 individuals with northern or Mexican grey wolf genotypes (Table 1). Specimens examined: 1974–76 samples from Texas (Chambers Co. [7], Brazoria Co. [27], Cameron Co. [3], Harris Co. [2], Jefferson Co. [12], Liberty Co. [2], Montague Co. [15], Webb Co. [6]) and Louisiana (Calcasieu Co. [3]), and historical samples from Texas (Parker Co., 1930 (NMNH 249694)), Louisiana, (Tallula, 1905 (NMNH 136106)), Arkansas, (Stillwater, 1921 (NMNH 236542)) and (Onyx, 1922 (NMNH 242564)), Oklahoma, (Red Fork, 1905 (NMNH 135747)), Missouri, (Barren, 1924 (NMNH 244246)). NE, Nebraska; KS, Kansas; OK, Oklahoma; TX, Texas; MO, Missouri; AR, Arkansas; LA, Louisiana; IA, Iowa; IL, Illinois; IN, Indiana; KY, Kentucky; TN, Tennessee; MS, Mississippi; AL, Alabama; OH, Ohio; WV, West Virginia; VA, Virginia; NC, North Carolina; SC, South Carolina; GA, Georgia; FL, Florida.

TABLE 2 Sequence comparisons

	Coyotes					Grey wolves			Dog	Red wolves					
	3	1	7	14, 24	22	1	4	Mex		Cap	Ark1	Ark2	Mo	La, Ok, Tx	Cau
CLA-3	—	0.3	0.8	1.3	1.0	5.8	6.0	5.5	6.8	0.2	0.2	1.0	0.1	5.8	7.3
CLA-1	1	—	1.0	1.0	1.3	6.0	6.1	5.8	7.0	1.0	1.5	1.2	1.0	6.0	7.0
CLA-7	3	4	—	0.5	0.3	5.5	5.8	5.3	6.5	1.0	1.5	0.5	0.5	5.5	7.5
CLA-14, 24	5	4	2	—	0.8	5.5	5.8	5.3	6.5	0.5	1.5	1.0	1.0	5.5	7.0
CLA-22	4	5	1	3	—	5.8	6.0	5.5	6.8	0.5	1.5	0.8	1.0	5.8	7.8
CLU-1	23	24	22	22	23	—	0.3	0.3	1.5	5.0	5.0	4.5	5.0	0.0	6.5
CLU-4	24	25	23	23	24	1	—	0.5	1.8	5.3	5.3	4.8	5.3	0.3	6.5
CLU-MEX	22	23	21	21	22	1	2	—	1.8	4.8	4.8	4.3	4.8	0.3	6.3
DOG	27	28	26	26	27	6	7	7	—	6.0	6.0	5.5	6.0	1.5	7.5
CRU-CAP	7	6	4	2	5	20	21	19	24	—	1.0	0.8	1.5	5.0	6.5
CRU-ARK1	7	6	6	6	7	20	21	19	24	4	—	0.8	1.5	5.0	7.0
CRU-ARK2	4	5	2	4	3	18	19	17	22	3	3	—	0.3	4.5	7.3
CRU-MO	3	4	2	4	3	20	21	19	24	6	6	1	—	5.0	7.0
CRU-LA, OK, TX	23	24	22	22	23	0	1	1	6	20	20	18	20	—	6.5
CAU	29	28	30	28	31	26	26	25	30	26	28	26	28	26	—

Percentage sequence divergence (above diagonal) and number of nucleotide substitutions (below diagonal) in 398 bp of the cytochrome *b* gene from samples of coyotes (CLA-), grey wolves (CLU-), captive (CRU-CAP) and historic red wolves (CRU-), a dog, and the golden jackal (CAU). See Fig. 1 for locality information. Methods. Two universal primers for cytochrome *b*, H1549 and L14725 (ref. 10), were synthesized and used to amplify 398 bp of mtDNA sequence. Double-stranded sequence was amplified first and used in a second PCR reaction to generate single-stranded template by the unbalanced primer method¹⁵. Each PCR reaction mixture contains ~10 ng genomic DNA; 10 mM dNTP mix in a reaction buffer of 50 mM KCl, 2.5 mM MgCl₂, 10 mM Tris-HCl (pH 8.3), and 2.5 U *Taq* DNA polymerase in 50 µl. Each primer (25 pmol) was used for the double-stranded amplification. For single-stranded amplification we used an unequal ratio of 25 to 0.25 pmol. Thirty-five to forty cycles of amplification were run in a programmable Perkin-Elmer Cetus DNA thermal cycler as follows: denaturation at 94 °C for 1 min, annealing at 50–55 °C for historical samples, or 60–65 °C for recent samples both for 1 min, extension at 72 °C for 1 min. Double-stranded reaction products (15 µl) were separated in a 3% Nusieve (FMC Corporation, Rockland, Maryland) agarose gel, the appropriate band was excised and resuspended in 10–100 µl H₂O. The double-stranded product (1–5 µl) was used to produce single-stranded template. After concentration with Centricon 100 microconcentrators (Amicon), 7 µl single-stranded product was sequenced using the limiting primer in the second PCR and a Sequenase kit (USB). Because of the sensitivity of PCR amplification, several controls were designed to detect contamination. Tissue obtained from museum samples was extracted using sterile materials and filtered pipette tips. Four historical samples were extracted in R.K.W.'s laboratory and two by S.M.J. at the University of California at San Francisco. All were then amplified and sequenced in a laboratory in which canine DNA had not previously been handled. As a control, tissues from two individuals were extracted in both laboratories, and the duplicate samples were sequenced. In each case the resulting sequence was identical in the duplicates. Finally, extraction controls and no-template PCR controls were also used.

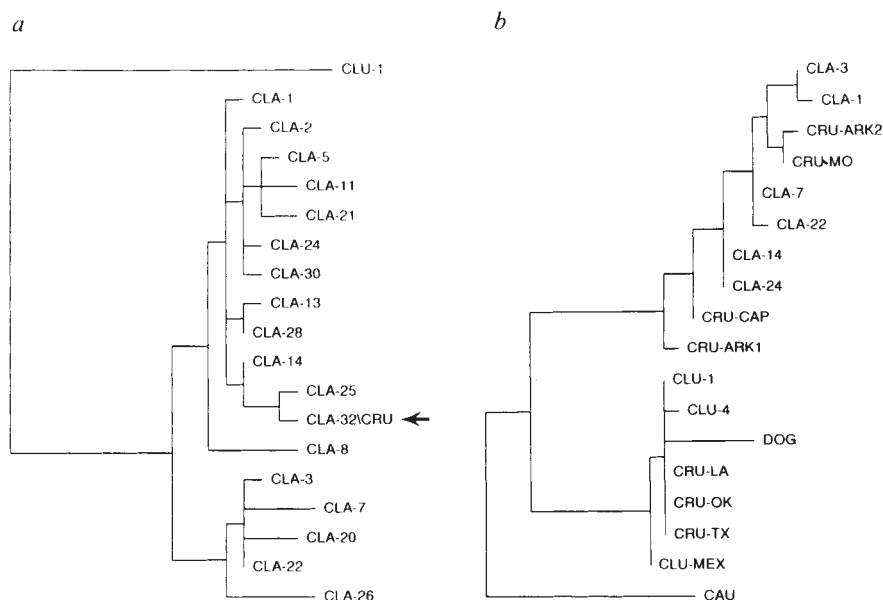
populations of coyotes⁸. The least our data show is that extensive hybridization has occurred among red wolves, coyotes and grey wolves.

Both empirical and theoretical studies indicate that a hybrid zone even as large as the past geographical range of the red wolf is possible^{7,8,13}. The present area of the hybrid zone between grey wolves and coyotes in Minnesota and Canada is about half the geographical range hypothesized for red wolves⁷. Furthermore, hybrid zone width can be 50 times the average dispersal distance¹³. Grey wolves and coyotes are large mobile predators and the distance moved in a single generation often exceeds 100 kilometres¹⁴. Thus with time, a hybrid zone several thousand

kilometres in width could develop. Further research, especially on nuclear genes, is needed to define better the temporal and spatial extent of the red wolf hybrid zone, but our results suggest that, relative to the past geographical range of the red wolf, it is the largest yet described for a terrestrial vertebrate.

The management policy of the US Fish and Wildlife Service does not grant protection to hybrids¹⁸. Nevertheless, protection of red wolves should be continued. Even if the red wolf is entirely a hybrid, it filled the role as top predator throughout its former geographic range and was thus an integral part of the ecosystem. The captive population of red wolves seems to be morphologically and genetically representative of the canid that

FIG. 2 *a*, A most parsimonious phylogenetic tree of coyote, *Canis latrans* (CLA), and captive red wolf (CRU) genotypes based on analysis of restriction site data^{7,8}. The arrow points to the single genotype found in a sample of eight red wolves representing the four known matrilineal lines in the captive population of red wolves. Genotype numbers correspond to those reported previously^{7,8}. The tree was generated using the branch-and-bound option of the PAUP program (version 2.4, D. Swofford, Illinois Natural History Survey, Champaign) and rooted by a genotype of the grey wolf, *Canis lupus* (CLU-1)⁷. Tree length, 70; overall consistency index, 0.85. *b*, The most parsimonious tree of recent and historical DNA samples of canids on the basis of 398 bp of sequence from cytochrome *b*. Species designations as above. Locality of historical specimens: MO, Missouri; ARK, Arkansas; LA, Louisiana; OK, Oklahoma; TX, Texas (see Fig. 1). Genotypes CLU-1 and CLU-4 are from wolves found in Minnesota and Alaska, respectively, and CLU-MEX is from a Mexican wolf. The PAUP-generated tree is rooted by sequence data from the golden jackal *Canis aureus* (CAU). Tree length, 56; overall consistency index, 0.86.



existed in the southeastern United States, and so its reintroduction there would restore an essential component of the fauna. Moreover, protection should continue unless loss of the hybrid form is judged to be reversible. The absence of red wolf-like individuals in the present hybrid zone in Minnesota and Canada suggests that unique environmental and genetic conditions are required for the genesis of a red wolf phenotype. □

Received 27 February; accepted 25 April 1991.

- Nowak, R. M. *North American Quaternary Canis* (Mongr. Mus. Nat. Hist. Univ. Kansas, Lawrence, 1979).
- McCarley, H. *Southwestern Naturalist* **7**, 227–235 (1962).
- Young, S. P. *The Wolves of North America* (Dover, New York, 1944).
- Lawrence, B. & Bossert, W. H. *Am. Zool.* **7**, 223–232 (1967).
- Ferrell, R. E. *et al. Biochem. Genet.* **18**, 39–49 (1980).
- Brown, W. M., George, M. & Wilson, A. C. *Proc. natn. Acad. Sci. U.S.A.* **76**, 1967–1971 (1979).
- Lehman, N. *et al. Evolution* **45**, 104–119 (1991).
- Lehman, N. & Wayne, R. K. *Genetics* (in the press).
- Wayne, R. K. *et al. Proc. natn. Acad. Sci. U.S.A.* **87**, 1772–1776 (1990).
- Meyer, A. & Wilson, A. C. *J. molec. Evol.* **31**, 359–370 (1990).
- Felsenstein, J. *Evolution* **39**, 783–791 (1985).
- Mech, L. D. *The Wolf: The Ecology and Behavior of an Endangered Species* (University of Minnesota, Minneapolis, 1970).
- Barton, N. H. & Hewitt, G. M. *Nature* **341**, 497–503 (1989).
- Mech, L. D. in *Mammalian Dispersal Patterns* (eds Chepko-Sade, B. D. & Halpin, Z. T.) 55–74 (University of Chicago Press, Chicago, 1987).
- Gyllenstein, U. B. & Erlich, H. A. *Proc. natn. Acad. Sci. U.S.A.* **85**, 7652–7656 (1988).
- Carbyn, L. N. *Gray Wolf and Red Wolf* (eds Novak, M., Baker, J. A., Obbard, M. E. & Malloch, B.) 358–377 (Ontario Ministry of Natural Resources, Toronto, 1987).

ACKNOWLEDGEMENTS. We thank R. Nowak, R. Smith and W. Parker for constructive comments and samples of red wolves; M. Carleton (US National Museum of Natural History) and R. Zink and J. Bates (Louisiana State University Museum of Natural Science) for their help; Y. F. C. Lau for providing facilities for PCR and DNA sequencing; A. Meyer, K. Chan, C. Nagamine and W. K. Thomas for technical advice; and B. Elder, E. Geffen, D. Girman, S. George, C. Hunter, P. Kat, N. Lehman, A. Mecure, L. D. Mech, S. J. O'Brien, U. Seal, N. Sturm and B. Van Valkenburgh for comments on the manuscript. Funding was provided by Point Defiance Zoo and Aquarium, the US Fish and Wildlife Service, and the NSF.

Disruption of retinogeniculate afferent segregation by antagonists to NMDA receptors

Jong-On Hahn, Ronald B. Langdon & Mriganka Sur*

Department of Brain and Cognitive Sciences,
Massachusetts Institute of Technology,
Cambridge, Massachusetts 02139, USA

AFFERENT activity has an important role in the formation of connections in the developing mammalian visual system^{1,2}. But the extent to which the activity of target neurons shapes patterns of afferent termination and synaptic contact is not known. In the ferret's visual pathway, retinal ganglion cell axons from each eye segregate early in development into eye-specific laminae in the lateral geniculate nucleus (LGN)³. The dorsal laminae (termed laminae A and A1) then segregate further into inner and outer sublaminae that retain input from on-centre and off-centre retinal axons, respectively^{4,5}. Thus, individual retinogeniculate axons form terminal arbors within laminae A and A1 that are restricted to one inner or outer sublamina⁶. We report here that blockade of *N*-methyl-D-aspartate (NMDA) receptors on LGN cells with specific antagonists during the period of sublamina formation prevents retinal afferents from segregating into 'On' and 'Off' sublaminae. Retinogeniculate axons have arbors that are not restricted appropriately, or are restricted in size but inappropriately positioned within the eye-specific laminae. NMDA receptor antagonists may specifically disrupt a mechanism by which LGN neurons detect correlated afferent and target activity⁷, and have been shown to reduce retinogeniculate transmission more generally^{8–10}, causing LGN cells to have markedly reduced levels of activity. These results therefore indicate that the activity of postsynaptic cells can significantly influence the patterning of

TABLE 1 Axons in normal and D-APV-treated animals

	No. of animals	No. of axons	Mean arbor area in μm^2 (s.e.)	Mean sublamina index (s.e.)
Normal 2-week-old	3	12	5250 (601)	0.70 (0.01)
Normal 3-week-old	3	16	8155 (653)	0.88 (0.05)
D-APV-treated, 3-week-old	3	12	7959 (807)	0.66 (0.03)

The arbors that were examined lay within the central two-thirds of the LGN, where the thickness of the A-laminae is roughly uniform, and their locations were judged to be between 0 and 60 degrees in azimuth and ± 30 degrees in elevation²⁴. In this region, in each animal, every arbor that could be fully reconstructed was drawn; we did not sample or select arbors (this procedure led to 2–7 reconstructed axons per animal in each group). The arbor reconstructions were made without knowledge of laminar borders or sublamina halves. The axons in each of the three groups spanned similar ranges of LGN locations. In each group, there were nearly equal numbers of arbors in the binocular and monocular portions of the LGN, and there were no differences between these axons. The area of an arbor was measured as the area covered by the outermost branch tips in the two-dimensional reconstruction of the arbor. The 'sublamina index' was obtained from the position of the arbor with respect to a line bisecting lamina A or A1 parallel to laminar borders. The bisected lamina halves closely approximate sublaminae; the sublamina index varies between 0.5 and 1, being 1 for an arbor that was entirely confined to one half of lamina A or A1 and 0.5 for an arbor that was located exactly midway between laminar borders. Statistical comparisons employed the Mann-Whitney *U*-test and were made by: (1) pooling the axons from each animal and treating each animal as a single datum, (2) treating each axon as a single datum.

inputs and the structure of presynaptic afferents during development.

Experiments were done on 32 ferret kits. Timed pregnant ferrets were either purchased from Marshall Research Animals or bred in our colony. We first examined, in normal ferret kits, the time course of segregation of retinogeniculate afferents into eye-specific laminae and 'On' and 'Off' sublaminae. At birth (embryonic day 41) axons from the two eyes overlap extensively in the LGN³. Axon arbors start to segregate into eye-specific laminae by the first postnatal week, and laminar segregation is essentially complete by two weeks (Fig. 1a; $n = 2$). Further division of laminae A and A1 into sublaminae occurs during the third postnatal week, and sublaminae can be clearly identified by three weeks (Fig. 1b; $n = 3$).

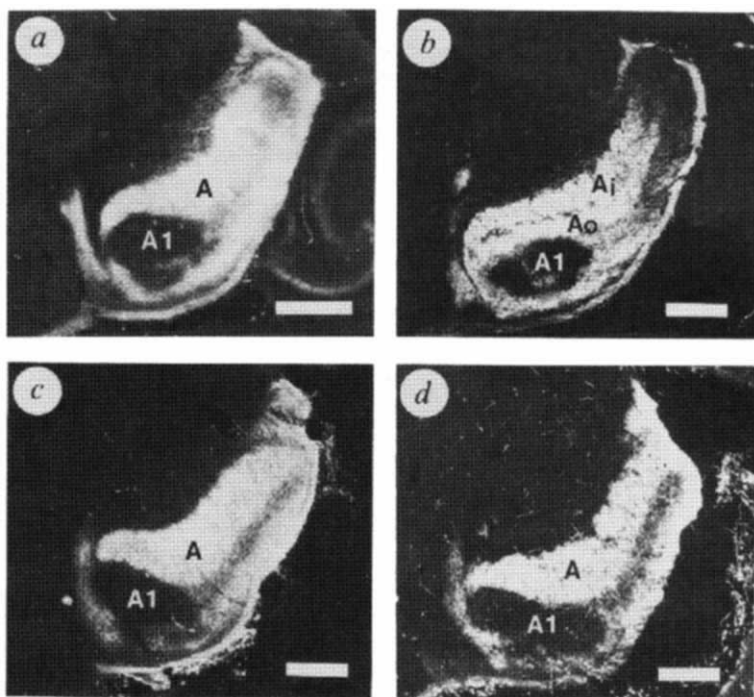
We next examined the effect on sublamina development of blocking NMDA receptors on LGN cells between two and three weeks of age. Antagonists were delivered using osmotic minipumps either directly into the thalamus, near the LGN, or systemically using subcutaneous infusion. We first assayed the effect of the antagonists using intraocular injections of the anterograde tracer wheat-germ agglutinin conjugated to horseradish peroxidase. Blockade of NMDA receptors in the LGN with D-APV (D-2-amino-5-phosphonopivalic acid) delivered into the posterior thalamus, or with MK-801 ((+)-5-methyl-10,11-dihydro-5H-dibenzocyclohepten-5,10-imine maleate) delivered subcutaneously, blocks the formation of On and Off sublaminae (Fig. 1c,d). These effects are dose-dependent. Although D-APV delivered at a minipump concentration of 0.8 mM ($n = 3$) blocks sublamina formation (Fig. 1c), a concentration of 0.08 mM does not ($n = 2$). MK-801 blocks sublamina formation at a concentration of 4.75 mM (Fig. 1d; $n = 2$) but not at 1.2 mM ($n = 2$). In contrast, thalamic infusion of 0.8 mM L-APV (the inactive isomer of APV) has no effect on sublaminae ($n = 2$). Similarly, thalamic ($n = 2$) or subcutaneous ($n = 1$) infusion of the saline vehicle has no effect on the formation of sublaminae.

We next examined the morphology of single retinal axon arbors in the LGN by filling optic tract axons with horseradish

* To whom correspondence should be addressed.

FIG. 1 Dark-field photomicrographs of horizontal sections from the right LGN in ferrets in which the left eye was injected with the tracer wheat-germ agglutinin conjugated to horseradish peroxidase (WGA-HRP) to label retinogeniculate terminations. *a*, LGN from a normal two-week-old animal, showing that projections from the contralateral eye to the dorsal layers of the LGN are restricted to lamina A, whereas lamina A1, which receives projections from the ipsilateral eye, is free of labelled terminations. *b*, LGN from a normal three-week-old animal, showing that the projections from the contralateral eye to lamina A have segregated further into inner (Ai) and outer (Ao) sublaminae. *c*, LGN from a three-week-old animal that received thalamic infusion of 0.8 mM D-APV, a competitive antagonist of NMDA receptors, from two weeks to three weeks of age. Sublaminae within lamina A, which would normally be present by this age, now fail to form. *d*, LGN from a three-week-old animal that received subcutaneous infusion of 4.75 mM MK-801, a noncompetitive antagonist of NMDA receptors, from two weeks to three weeks of age. Sublaminae fail to form within lamina A. Bars, 250 μ m.

METHODS. Osmotic minipumps (Alzet model 2001, infusion rate $1 \mu\text{l h}^{-1}$) were implanted into two-week-old ferret kits. Each minipump contained one of the following (in saline): 0.8 mM and 0.08 mM D-APV, 0.8 mM L-APV, 4.75 mM and 1.2 mM MK-801, as well as isotonic saline. D-APV, L-APV and saline were infused into the thalamus, whereas MK-801 and saline were infused subcutaneously. For thalamic infusion, each minipump was connected via a catheter (a piece of polyvinyl tubing 1.5 cm long) to a 7-mm-long 28-gauge stainless steel cannula (Plastics One). After anaesthesia with ketamine (4 mg per 100 gm body weight) and metofane (1–2%), the skin over the skull was incised and separated from subcutaneous tissue at the base of the neck to create a pocket for the minipump. A small hole was drilled in the skull, the dura punctured and the cannula inserted into the brain with the tip resting in the thalamus rostral and medial to the LGN. The cannula was glued to the skull with cyanoacrylate glue and cemented with dental acrylic. Subcutaneous infusion was done by inserting a minipump near the base of the skull. The skin was sutured and the animal returned to the mother. After six days, the animal received an injection of WGA-HRP in the eye contralateral to the thalamic implant. After another 24 h of



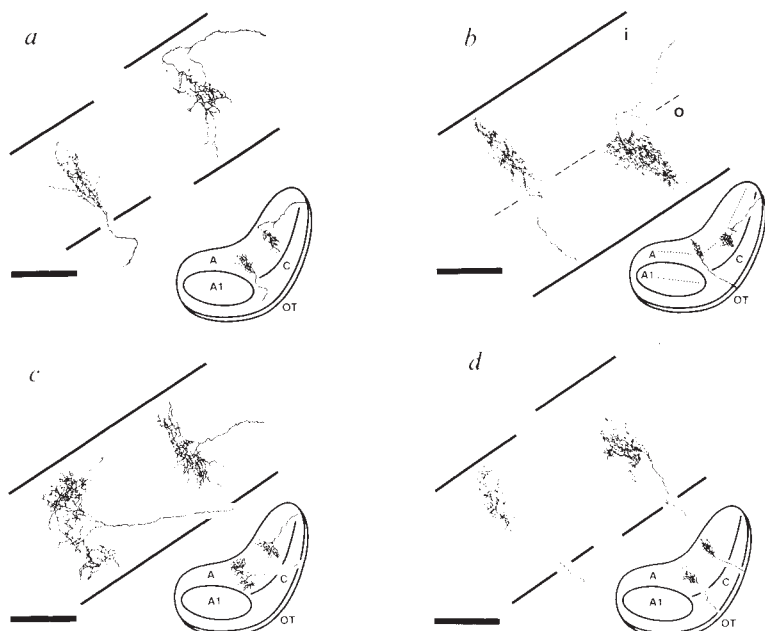
survival, animals were perfused with aldehydes, the brain cut at 50 μ m in the horizontal plane and sections reacted with tetramethyl benzidine for visualizing the HRP reaction product. In animals treated with D-APV, we estimate the concentration of D-APV in the LGN as 5–10 μ M, based on the distance of the cannula tip from the centre of the LGN and the dilution curve obtained by Bear *et al.*¹⁵. This concentration is adequate to selectively attenuate the NMDA component of optic tract evoked excitatory postsynaptic potentials recorded intracellularly in slices of the LGN from adult and developing ferrets (refs 22, 23; M. Esguerra, C. A. White and M.S., unpublished observations).

peroxidase *in vitro*. In normal ferret kits at two weeks of age ($n=3$), individual retinogeniculate axon arbors span one eye-specific lamina (Fig. 2*a*). By three weeks ($n=3$), when On and Off sublaminae have formed, individual arbors are restricted to one sublamina (Fig. 2*b*). Application of D-APV during the third postnatal week ($n=3$) alters the morphology of individual axon

arbors, leading to arbors that differ systematically from normal three-week arbors (Fig. 2*c, d*). At one extreme, some arbors span the height of an eye-specific lamina and fail to restrict appropriately into On and Off sublaminae (Fig. 2*c*). At the other extreme, arbors restrict in size such that they would fit in sublaminae, and some are appropriately positioned at the inner or

FIG. 2 Horseradish peroxidase-labelled retinal ganglion cell axon arbors within the LGN of normal ferrets and ferrets treated with D-APV. *a*, Axons in a normal two-week-old animal; arbors span nearly the entire extent of lamina A (or, not shown, lamina A1). *b*, Axons in a normal three-week-old animal; arbors are restricted to an inner (i) or outer (o) sublamina within lamina A (or, not shown, lamina A1). Dashed lines mark sublaminar borders. *c, d*, Axons in three-week-old animals treated with 0.8 mM D-APV from two weeks to three weeks of age. Some axons (shown in *c*) span the height of lamina A. Other axons (shown in *d*) are restricted in size but are positioned either normally, at the inner or outer half of lamina A, or abnormally, at the center of lamina A. Bars, 100 μ m and apply to large-scale axon drawings.

METHODS. Minipumps filled with 0.8 mM D-APV were implanted into two-week-old ferret kits as described in the legend to Fig. 1. After one week of survival, animals were deeply anaesthetized with sodium pentobarbital and transcardially perfused with cold (4 °C) artificial cerebrospinal fluid. The brain was quickly removed, the cortex dissected away, and the thalamus exposed and bisected. Small deposits of horseradish peroxidase were placed in the optic tract ventral to the LGN. The tissue was maintained in artificial cerebrospinal fluid at room temperature for 4–6 h to allow transport of horseradish peroxidase to axon terminals, then fixed by immersion and sectioned at 100 μ m in the horizontal plane. Sections were reacted with diaminobenzidine for visualizing horseradish peroxidase. Axon arbors were reconstructed from serial sections under a camera lucida, using a $\times 63$ objective for transmitted light and a $\times 100$ objective for phase contrast microscopy.



outer half of an eye-specific lamina, as in normal animals (Fig. 2d). Many of the restricted arbors, however, are inappropriately positioned in the eye-specific laminae, being located towards the centre of the lamina rather than in the inner or outer half (Fig. 2d).

In normal animals, arbor areas increase between two and three weeks of age (Table 1, $P < 0.05$, Mann-Whitney U -test, treating each animal as a single datum; $P < 0.01$ treating each axon as a single datum). In animals treated with D-APV, arbor sizes are intermediate between two- and three-week-old normal animals. We defined a 'sublamina index' as the proportion of an arbor's area lying in the half (either inner or outer) of lamina A or A1 containing the greatest portion of that arbor. A sublamina index of 0.5 thus represents a total absence of sublamination (equal division of an arbor between inner and outer halves), and 1.0 signifies complete restriction to a single sublamina. The index increases between two and three weeks in normal animals (Table 1, $P < 0.05$ comparing individual animals; $P < 0.01$ comparing individual axons), consistent with the segregation of arbors into On and Off sublaminae during this period. In D-APV treated animals, the index is similar to normal two-week-old animals but lower than in normal three-week-old animals ($P < 0.05$ comparing individual animals; $P < 0.005$ comparing individual axons), consistent with arbors that vary from being less restricted in size with respect to the A-laminae to being restricted but located inappropriately within the laminae.

The effects of NMDA receptor blockade on the development of axon arbors have not been examined previously in the mammalian brain, or in any other system in which the receptor is involved directly in synaptic transmission and in regulating the level of postsynaptic neuronal activity. In amphibia, NMDA receptor antagonists disrupt the segregation of eye-specific stripes in experimentally created retinal projections to the tectum¹¹. Individual retinotectal afferents in these animals have arbors that are more dispersed than normal¹² and have reduced branch density¹³. Physiological effects of NMDA receptor blockade during development include disruption of the alignment of maps from the two eyes in the amphibian tectum¹⁴, and prevention of ocular dominance plasticity in the kitten visual cortex¹⁵ (see, however, refs 16–19). Infusion of tetrodotoxin into the thalamus alters the morphology of retinogeniculate terminations in fetal cats^{20,21}, indicating a role for activity-dependent mechanisms in shaping afferent arbors. Tetrodotoxin, however, acts by blocking action potentials, and would do so both pre- and postsynaptically when infused intrathalamically.

NMDA receptor antagonists in the developing ferret LGN may specifically disrupt a mechanism detecting temporal correlations between retinal axon and LGN cell activity⁷. More generally, blockade of NMDA receptors during development reduces retinogeniculate transmission and the activity of postsynaptic LGN cells. Iontophoresis of NMDA antagonists in adult cats severely reduces the visual and background responses of LGN cells^{8–10}; NMDA receptors contribute substantially to the optic tract evoked excitatory postsynaptic potential in cells recorded *in vitro* from the LGN of adult^{22,23} and developing ferrets (M. Esguerra, C. A. White and M.S., unpublished observations). Thus the effects we observe on retinogeniculate axon arbors suggest strongly that the structure of retinal afferents in general, and the restriction of retinal afferents into On and Off sublaminae in the ferret LGN in particular, is regulated significantly by the activity of LGN cells during development.

9. Heggelund, P. & Hartveit, E. *J. Neurophysiol.* **63**, 1347–1360 (1990).
10. Kwon, Y. H., Esguerra, M. & Sur, M. *J. Neurophysiol.* (in the press).
11. Cline, H. T., Debski, E. & Constantine-Paton, M. *Proc. natn. Acad. Sci. U.S.A.* **84**, 4342–4345 (1987).
12. Cline, H. T. & Constantine-Paton, M. *Neuron* **3**, 413–426 (1989).
13. Cline, H. T. & Constantine-Paton, M. *J. Neurosci.* **10**, 1197–1216 (1990).
14. Scherer, W. J. & Udin, S. B. *J. Neurosci.* **9**, 3837–3843 (1989).
15. Bear, M. F., Kleinschmidt, A., Gu, Q. & Singer, W. *J. Neurosci.* **10**, 909–925 (1990).
16. Reiter, H. O. & Stryker, M. P. *Proc. natn. Acad. Sci. U.S.A.* **85**, 3623–3627 (1988).
17. Miller, K. D., Chapman, B. & Stryker, M. P. *Proc. natn. Acad. Sci. U.S.A.* **86**, 5183–5187 (1989).
18. Fox, K., Sato, H. & Daw, N. *J. Neurosci.* **9**, 2443–2454 (1989).
19. Langdon, R. B. & Sur, M. *J. Neurophysiol.* **64**, 1484–1501 (1990).
20. Shatz, C. J. & Stryker, M. P. *Science* **242**, 87–89 (1988).
21. Sretavan, D. W., Shatz, C. J. & Stryker, M. P. *Nature* **335**, 468–471 (1988).
22. Esguerra, M. & Sur, M. *Neurosci. Abstr.* **16**, 159 (1990).
23. Esguerra, M., Kwon, Y. H. & Sur, M. *Neurosci. Abstr.* **15**, 175 (1989).
24. Zahs, K. R. & Stryker, M. P. *J. comp. Neurol.* **241**, 210 (1985).

ACKNOWLEDGEMENTS. We thank T. Sullivan for technical assistance, C. White, S. Pallas, A. Ramoa and A. Roe for comments on the manuscript. This work was supported by the NIH and the Whitaker Fund.

Somatostatin stimulates Ca^{2+} -activated K^+ channels through protein dephosphorylation

Richard E. White*, Agnes Schonbrunn† & David L. Armstrong*‡

* Laboratory of Cellular and Molecular Pharmacology, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina 27709, USA

† Department of Pharmacology, University of Texas Medical School, Houston, Texas 77225, USA

THE neuropeptide somatostatin inhibits secretion from electrically excitable cells in the pituitary, pancreas, gut and brain¹. In mammalian pituitary tumour cells somatostatin inhibits secretion through two distinct pertussis toxin-sensitive mechanisms^{2–5}. One involves inhibition of adenyl cyclase⁶, the other an unidentified cyclic AMP-independent mechanism that reduces Ca^{2+} influx^{7,8} by increasing membrane conductance to potassium^{9,10}. Here we demonstrate that the predominant electrophysiological effect of somatostatin on metabolically intact pituitary tumour cells is a large, sustained increase in the activity of the large-conductance Ca^{2+} - and voltage-activated K^+ channels (BK). This action of somatostatin does not involve direct effects of Ca^{2+} , cAMP or G proteins on the channels. Our results indicate instead that somatostatin stimulates BK channel activity through protein dephosphorylation.

Previous electrophysiological studies of rat pituitary tumour cells have demonstrated that somatostatin reduces voltage-activated Ca^{2+} currents and increases inwardly rectifying K^+ currents^{11–13}. Because these effects are blocked by pertussis toxin but persist in the presence of exogenous cAMP, they are believed to be mediated by G proteins interacting directly with the channels^{14–16}. However, in previous studies cells were dialysed internally with EGTA-buffered salt solutions, conditions under which the responses to somatostatin are small and run down gradually, even though GTP is added to the dialysis solution^{11–13}. To avoid these problems, we voltage-clamped GH_4C_1 cells through nystatin-permeabilized membrane patches^{17,18}. With this new method, the cytoplasmic concentration of divalent cations and small metabolites is not disturbed, and membrane currents remain stable for over an hour¹⁸.

The pharmacologically isolated Ca^{2+} current elicited by membrane depolarization from -40 mV (Fig. 1a) passes almost exclusively through dihydropyridine-sensitive L-type channels¹⁹. Bath application of somatostatin inhibited this Ca^{2+} current at all voltages (Fig. 1b). But the peak current was reduced by only $28 \pm 3\%$ ($n = 5$) by maximally effective concentrations (≥ 100 nM) of somatostatin. In contrast, dihydropyridines must

Received 27 December 1990; accepted 12 April 1991.

1. Movshon, J. A. & Van Sluyters, R. C. *A. Rev. Psychol.* **32**, 477–522 (1981).
2. Sherman, S. M. & Spear, P. D. *Physiol. Rev.* **62**, 740–855 (1982).
3. Linden, D. C., Guillery, R. W. & Cucchiari, J. *J. comp. Neurol.* **203**, 189–211 (1981).
4. Stryker, M. P. & Zahs, K. R. *J. Neurosci.* **3**, 1943–1951 (1983).
5. Hahn, J. & Sur, M. *Neurosci. Abstr.* **14**, 460 (1988).
6. Roe, A. W., Garraghty, P. E. & Sur, M. *J. comp. Neurol.* **288**, 208–242 (1989).
7. Constantine-Paton, M., Cline, H. T. & Debski, E. *A. Rev. Neurosci.* **13**, 129–134 (1990).
8. Sillito, A. M., Murphy, P. C., Salt, T. E. & Moody, C. I. *J. Neurophysiol.* **63**, 347–355 (1990).

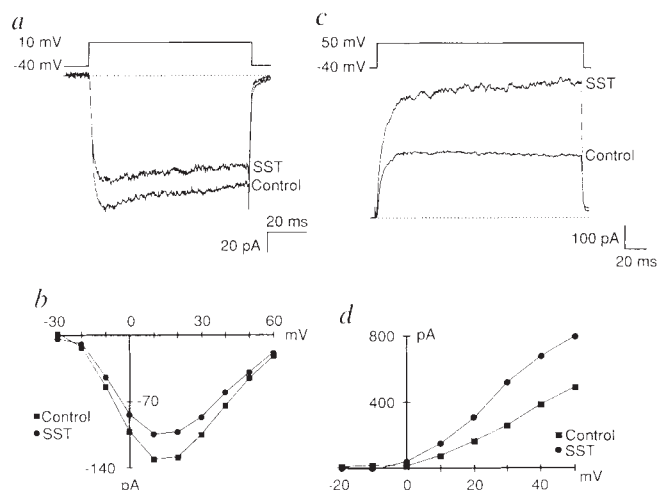
‡ To whom correspondence should be addressed.

FIG. 1 Somatostatin (SST) modulates voltage-activated currents in metabolically intact pituitary tumour cells. *a*, Voltage-activated inward current through dihydropyridine-sensitive Ca^{2+} channels is reduced by exposure to 100 nM somatostatin. *b*, Current-voltage relation of the peak dihydropyridine-sensitive current before and after the addition of somatostatin. *c*, Voltage-activated outward current is increased by 100 nM somatostatin. *d*, Current-voltage relation of the steady-state outward current before and after the addition of somatostatin.

METHODS. GH_4C_1 cells³⁹ were maintained as described previously¹⁹, and voltage-clamped to -40 mV at room temperature through nystatin-perforated patches¹⁸. The bath solution contained 140 mM NaCl, 5 mM KCl, 2 mM MgCl_2 , 5 mM CaCl_2 , 1 μM tetrodotoxin, 20 mM glucose and 20 mM HEPES at pH 7.2. Patch pipettes were manufactured from capillaries of Corning glass 7052. The tip of the patch pipette was filled with a solution containing 140 mM KCl, 5 mM MgCl_2 and 40 mM HEPES at pH 7.2. (To isolate voltage-activated Ca^{2+} channels, KCl in the pipette was replaced by CsCl, and 20 mM tetraethylammonium ions were added to the bath.) The remainder of the pipette was back-filled with a similar solution to which 200 $\mu\text{g ml}^{-1}$ nystatin (diluted by sonication from a 50 mg ml^{-1} stock in dimethylsulphoxide) had been added. Cells were studied further only if the voltage error across the series resistance could be reduced to ≤ 5 mV within 10–20 min after forming a gigaohm seal. Voltage-activated currents were filtered at 2 kHz and digitized at 10 kHz. Capacitive and leakage currents were subtracted digitally. Somatostatin was diluted to its final concentration in the bath solution and perfused into the 0.5-ml bath at 2–3 ml min^{-1} .

be applied at concentrations that completely block the Ca^{2+} current to inhibit secretion as effectively as somatostatin^{10,20}. Furthermore, experiments using fluorescent indicators in intact cells have shown that changes in K^+ conductance are responsible for the inhibition of Ca^{2+} entry and secretion by somatostatin^{9,10}. We therefore investigated the action of somatostatin on K^+ currents.

When the patch pipette contained 140 mM KCl instead of CsCl, the same voltage steps in physiological saline elicited a slowly activating, but sustained, outward current that completely obscured the underlying Ca^{2+} current (Fig. 1c). Maximally effective concentrations of somatostatin stimulated this potassium current at all voltages ($98 \pm 12\%$ at $+40$ mV; $n = 10$) without shifting the threshold of activation (Fig. 1d). The response to somatostatin required 5–10 minutes to reach maximum amplitude, but in the absence of further pharmacological manipulations, the outward current remained fully elevated for over an hour, even after washing out the somatostatin. Our observation that blocking the Ca^{2+} current with 2 mM CoCl_2 also decreased the outward current ($40 \pm 15\%$ at



Whenever experimental conditions eliminated the response to somatostatin, positive controls were carried out on untreated coverslips from the same plating. Average values are reported routinely as mean $\pm$ standard deviation, with the number of cells (n) indicated in parentheses.

$+30$ mV) and blocked the effect of somatostatin ($n = 4/4$) suggested that a Ca^{2+} -activated K^+ channel might be the target of somatostatin action.

Like other secretory cells, rat pituitary tumour cells express a class of large-conductance K^+ channels (BK) that are activated by both Ca^{2+} and voltage^{21–23}. As expected, single channel records from GH_4C_1 cells revealed a prominent class of large unitary K^+ currents (Fig. 2a). In cell-free patches these outwardly rectifying channels exhibited a conductance of 120 pS in physiological K^+ and ~ 180 pS in equimolar KCl (Fig. 2b). Channel activity in cell-free patches was unaffected by 2 mM Co^{2+} (not shown) but was blocked completely by 30 nM charybdotoxin²⁴ or 1 mM tetraethylammonium ions on the 'extracellular' side of the patch (Fig. 2a). The same maximal concentration of tetraethylammonium ions reduced the control outward current by $80 \pm 8\%$ on average ($n = 8$), and completely reversed the effect of somatostatin (Fig. 2c; $n = 3/3$). Furthermore, somatostatin had no effect ($n = 0/5$) on the residual outward current ($33 \pm 4\%$ of control at $+30$ mV) when 30 nM charybdotoxin had been added previously (Fig. 2d).

FIG. 2 Somatostatin (SST) increases outward current through charybdotoxin (CTX)- and tetraethylammonium (TEA)-sensitive potassium channels (BK) with a unitary conductance of 120 pS in physiological K^+ gradients. *a*, Representative records of unitary potassium channel activity at $+30$ mV from the same cell-free patch in the outside-out configuration. 30 nM CTX and 1 mM TEA selectively inhibit the activity of a 120 pS channel. *b*, Single channel current-voltage relation from outside-out patches exposed to physiological (■) or symmetrical (●) gradients of potassium. The data represent the average values from 4–7 patches. *c*, 1 mM TEA completely blocks the outward current stimulated by 100 nM SST at $+40$ mV. *d*, 30 nM CTX reduces steady-state outward current at $+40$ mV by more than 50% and completely blocks the response to 100 nM SST.

METHODS. Cell-free patches in the outside-out configuration were formed by omitting nystatin and adding 1 mM EGTA or 5 mM BAPTA ($\text{pCa} = 7$) to the solution in the patch pipette. After forming a gigaohm seal, the patch was disrupted by suction and excised from the cell.

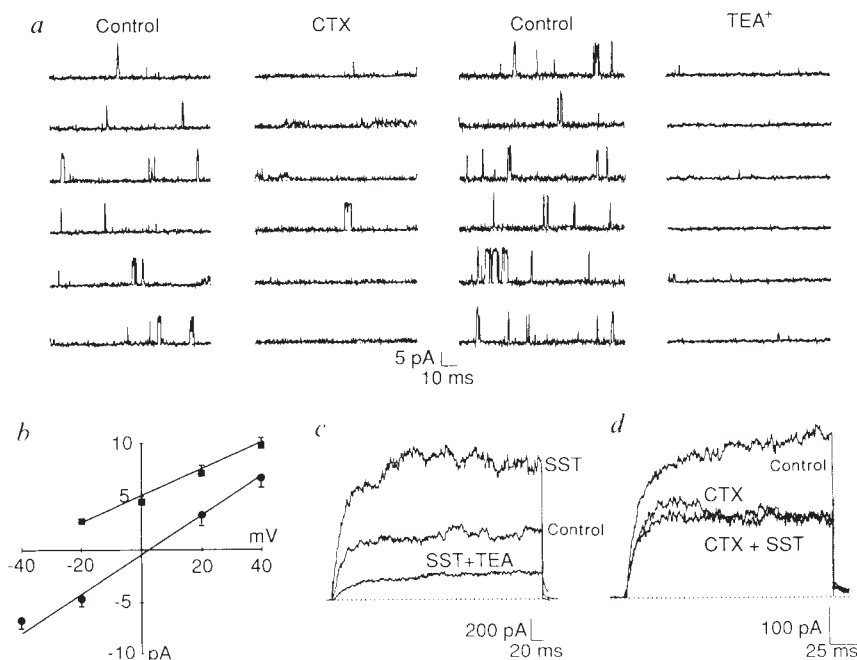
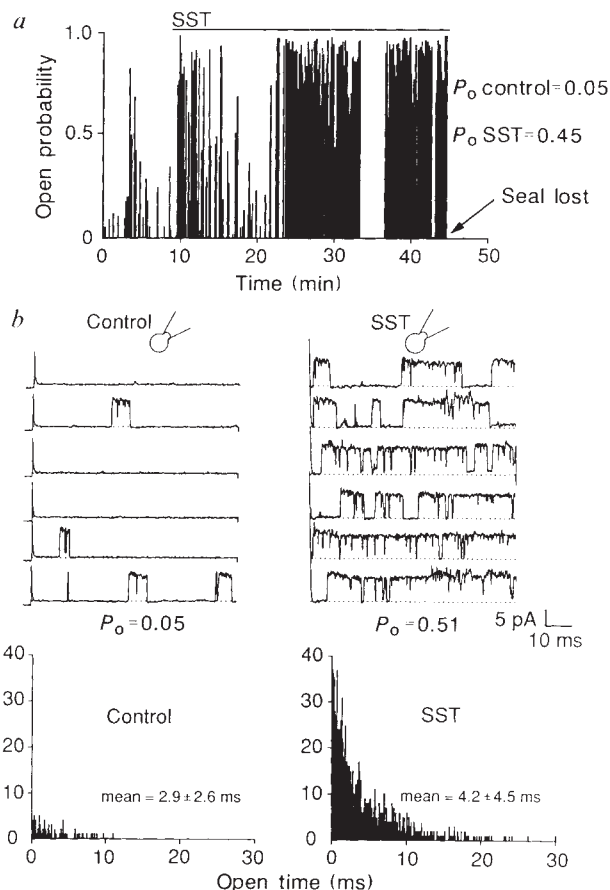


FIG. 3 Somatostatin (SST) increases the activity of BK channels in cell-attached patches on intact cells. *a*, Channel activity versus time in a patch with only one active channel. SST (100 nM) was added to the bath after 9 min, and increased the probability of finding the channel in the open state (P_o) from 0.05 to 0.45 over 35 min. *b*, Representative records of unitary K^+ currents from another cell-attached patch in the absence (control) and presence (SST) of 100 nM somatostatin. This patch also contained only one active BK channel, and SST increased P_o from 0.05 to 0.51. *c*, Frequency histograms of channel open times from the same patch in *b*. **METHODS.** Cell-attached patches were formed by filling the patch pipette with the standard bath solution minus glucose and making a gigaohm seal on an intact cell. The membrane potential across the rest of the cell membrane outside the patch pipette was eliminated by bathing the cell in a high-concentration K^+ solution which contained 140 mM KCl, 10 mM $MgCl_2$, 0.1 mM $CaCl_2$, 10 mM HEPES and 30 mM glucose (pH 7.2). Channel activity is plotted as the probability (P_o) of finding any channel in the open state during a 100-ms step to +40 mV. The steps were repeated at 5-s intervals, and the values of P_o are plotted as vertical lines versus time after the beginning of the recording.

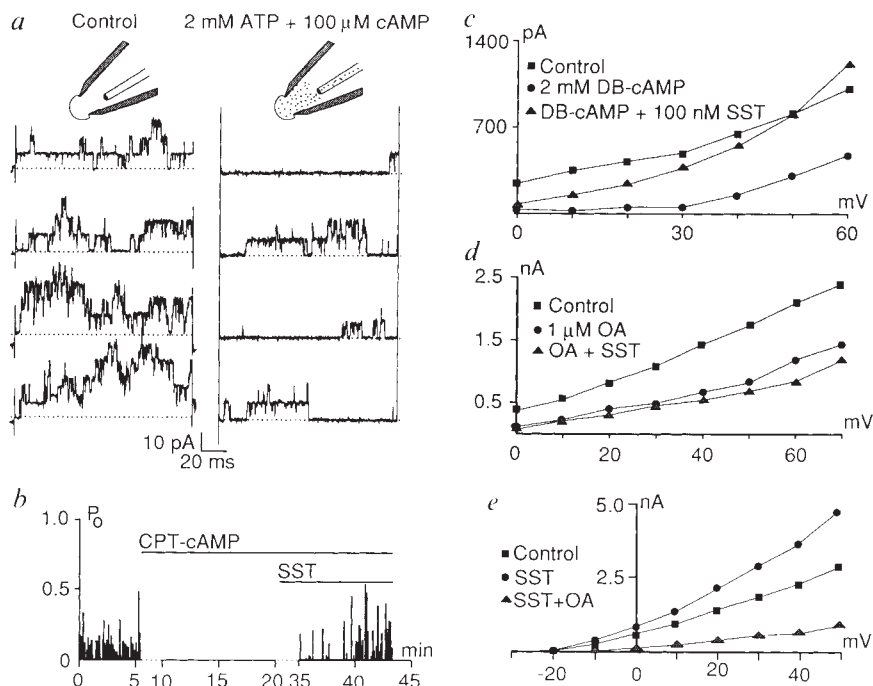


Direct evidence that BK channels are stimulated by somatostatin is presented in Fig. 3. In cell-attached patches with only one active BK channel, adding somatostatin to the bath outside the pipette rapidly increased the fraction of time the channel spent in the open state (Fig. 3*a, b*). Frequency histograms of channel open times (Fig. 3*c*) revealed no significant effect of somatostatin on the average duration of each opening. In contrast, the fraction of time the channel spent in the open state increased ten times from 0.05 to 0.49. Furthermore, in 17 cell-attached patches with more than one BK channel (data not shown), somatostatin increased the maximum number of simultaneously active channels from 1.9 ± 0.8 to 3.3 ± 1.4 ($P < 10^{-4}$, paired *t*-test) and increased $N \cdot P_o$ from 0.38 to 1.15 ($P < 0.02$, paired *t*-test). In summary, the predominant effect of somatostatin on membrane excitability of metabolically intact pituitary tumour cells is a large increase in the opening frequency of large-conductance, charybdotoxin- and tetraethylammonium-sensitive, Ca^{2+} - and voltage-activated K^+ channels. This effect of somatostatin might have gone unnoticed in earlier studies with conventional whole-cell recording because the Ca^{2+} buffers

that are added routinely to the pipette solution dialysing the cell interior would have suppressed the activity of BK channels.

The effect of somatostatin on BK channel activity, like its effects on secretion^{4,5}, adenylyl cyclase⁶ and ionic currents¹⁴⁻¹⁶, was blocked by pre-exposing the cells overnight to 100 ng ml⁻¹ pertussis toxin ($n = 5/6$; data not shown). Although this result indicates that a G protein couples somatostatin binding to BK

FIG. 4 BK channel activity is regulated by cAMP-dependent protein phosphorylation. *a*, Representative records of unitary potassium currents from a cell-free patch in the outside-out configuration. Perfusion of the pipette interior with ATP and cAMP dramatically reduced the activity of the large conductance channels. The maximal number of simultaneously active BK channels decreased from 5 to 2. *b*, BK channel activity versus time in a cell-attached patch. The cAMP analogue, 8-chlorophenylthio-cAMP (CPT-cAMP; 0.1 mM), and somatostatin (100 nM) were applied to the bath during the times indicated. To conserve space on the computer's hard disk, the recording was interrupted at the times indicated by dots on the time axis. *c*, Current-voltage relation of steady-state outward current obtained with the nystatin technique. Somatostatin (100 nM) overcomes the inhibition produced by bath application of 2 mM dibutyl-cAMP (DB-cAMP). *d*, Okadaic acid inhibits the current-voltage relation of steady-state outward current obtained with the nystatin technique. Somatostatin (100 nM) cannot overcome the inhibition produced by 1 μ M okadaic acid. *e*, Okadaic acid (OA; 1 μ M) reverses somatostatin-induced (100 nM) stimulation of steady-state outward current obtained with the nystatin technique.



channel activity²⁵, it is unlikely that the G protein activates the channels directly^{26,27}. Adding somatostatin to the bath rapidly stimulated channels isolated in cell-attached patches. In addition, we were unable to reconstitute the effect of somatostatin in cell-free patches by adding 100 μ M GTP to the solution on the cytoplasmic side of the membrane ($n = 0/6$, data not shown). Thus, G-protein activation by somatostatin seems to produce a readily diffusible second messenger which increases BK channel activity. This intracellular messenger is unlikely to be calcium²⁸ because somatostatin reduces Ca^{2+} entry through voltage-activated channels (Fig. 1a, b) and does not release calcium from internal stores when the membrane potential is clamped below the threshold for channel activation⁹.

The intracellular messenger is also unlikely to be cAMP. Although BK channels in other preparations are stimulated by cAMP-dependent protein phosphorylation^{29–31}, BK channels in primary rat gonadotrophs³² and GH_4C_1 cells (Fig. 4a) are inhibited. Therefore, somatostatin could stimulate BK channels by inhibiting adenylyl cyclase; however, even in the presence of saturating concentrations of exogenous cAMP analogues, somatostatin stimulated BK channel activity in cell-attached patches (Fig. 4b; $n = 3/5$) and intact cells (Fig. 4c; $n = 4/7$). Because somatostatin has no effect on phosphodiesterase activity in GH_4C_1 cells², these results indicate that somatostatin reverses the effects of cAMP-dependent protein phosphorylation on BK channels at a step distal to cAMP production and degradation. In principle, somatostatin could reverse the action of cAMP directly by noncompetitive inhibition of the kinase, or indirectly by activation of a protein phosphatase. In either case, blocking protein phosphatases should reduce the effectiveness of somatostatin.

Okadaic acid is a potent and highly specific inhibitor of serine/threonine protein phosphatases 1 and 2A (refs 33, 34). Like membrane-permeable derivatives of cAMP, okadaic acid by itself significantly reduced steady-state outward current ($80 \pm 9\%$ inhibition; $n = 3$) at a concentration (10 nM) that selectively inhibits phosphatase 2A *in vitro*³³. In contrast, 1 μ M okadaic acid had no effect on BK channel activity in cell-free patches when ATP was absent ($n = 0/2$, not shown). Unlike the effects of cAMP derivatives, the effect of 1 μ M okadaic acid on intact cells was not reversed by somatostatin (Fig. 4d; $n = 0/5$). Furthermore, 1 μ M okadaic acid completely reversed the effect of somatostatin on intact cells when it was applied after the neuropeptide (Fig. 4e, $n = 2$).

We conclude from these findings that somatostatin stimulates BK channels through dephosphorylation of a cAMP-dependent phosphorylation site on the channel protein or a closely associated regulatory molecule. Protein dephosphorylation could also account for the direct inhibitory effect of somatostatin on dihydropyridine-sensitive Ca^{2+} channels (Fig. 1a)^{11–15}, which require cAMP-dependent phosphorylation to respond to membrane depolarization^{35,36}. Whether dephosphorylation results from inhibition of the cAMP-dependent kinase or from activation of a protein phosphatase remains to be determined. But, the activity of one class of charybdotoxin-sensitive BK channels reconstituted from mammalian brain into lipid bilayers can be stimulated selectively by phosphatase 2A (ref. 37). In addition, activation of a protein phosphatase provides a simpler explanation for the observation that somatostatin is equally effective at inhibiting secretion whether it is stimulated through adenylyl cyclase or phospholipase $\text{C}^{4,5}$. In fact somatostatin has been reported to stimulate tyrosine phosphatase activity in pancreatic tumour cells³⁸. Together, these findings raise the possibility that all cAMP-independent effects of somatostatin are mediated by one or more protein phosphatases. □

- Koch, B., Dorflinger, L. J. & Schonbrunn, A. *J. biol. Chem.* **260**, 13138–13145 (1985).
- Yajima, Y., Akita, Y. & Saito, T. *J. biol. Chem.* **261**, 2684–2689 (1986).
- Koch, B. & Schonbrunn, A. *Endocrinology* **114**, 1784–1790 (1984).
- Schlegel, W., Wuarin, F., Wollheim, C. B. & Zahnd, G. R. *Cell Calcium* **5**, 223–236 (1984).
- Schlegel, W. *et al. Nature* **329**, 719–721 (1987).
- Koch, B. D., Blalock, J. B. & Schonbrunn, A. *J. biol. Chem.* **263**, 216–225 (1988).
- Koch, B. D. & Schonbrunn, A. *J. biol. Chem.* **263**, 226–234 (1988).
- Luini, A., Lewis, D., Guild, S., Schofield, G. & Weight, F. J. *Neurosci.* **6**, 3128–3132 (1986).
- Yamashita, N., Shibuya, N. & Ogata, E. *Proc. natn. Acad. Sci. U.S.A.* **83**, 6198–6202 (1986).
- Mollard, P., Vacher, P., Dufy, B. & Barker, J. L. *Endocrinology* **123**, 721–732 (1988).
- Lewis, D. L., Weight, F. F. & Luini, A. *Proc. natn. Acad. Sci. U.S.A.* **83**, 9035–9039 (1986).
- Rosenthal, W. *et al. EMBO J.* **7**, 1627–1633 (1988).
- Yamashita, N., Shibuya, N. & Ogata, E. *Proc. natn. Acad. Sci. U.S.A.* **85**, 4924–4928 (1988).
- Horn, R. & Marty, A. *J. gen. Physiol.* **92**, 145–159 (1988).
- Korn, S. J. & Horn, R. *J. gen. Physiol.* **94**, 789–812 (1989).
- Kalman, D., O'Laigue, P., Erleben, C. & Armstrong, D. *J. gen. Physiol.* **92**, 531–548 (1988).
- Enyeart, J. J., Aizawa, T. & Hinkle, P. M. *Am. J. Physiol.* **248**, C510–C519 (1985).
- Dubinsky, J. M. & Oxford, G. S. *J. gen. Physiol.* **83**, 309–339 (1984).
- Ritchie, A. K. *J. Physiol. (Lond.)* **385**, 591–609 (1987).
- Latorre, R., Oberhauser, A., Labarca, P. & Alvarez, O. A. *Rev. Physiol.* **51**, 385–399 (1989).
- Miller, C., Moczydlowski, E., Latorre, R. & Phillips, M. *Nature* **313**, 316–318 (1988).
- Pfeuffer, T. & Helmreich, E. J. M. *Curr. Top. cell. Reg.* **29**, 129–216 (1988).
- Yatani, A., Codina, J., Sekura, R. D., Birnbaumer, L. & Brown, A. M. *Molec. Endocr.* **1**, 283–289 (1987).
- Toro, L., Ramos-Franco, J. & Stefani, E. *J. gen. Physiol.* **96**, 373–394 (1990).
- Maruyama, Y. & Petersen, O. H. *Nature* **300**, 61–63 (1982).
- Ewald, D. A., Williams, A. & Levitan, I. B. *Nature* **315**, 503–506 (1985).
- Lechleiter, J. D., Dartt, D. A. & Brehm, P. *Neuron* **1**, 27–235 (1988).
- Kume, H., Takai, A., Tokumo, H. & Tomita, T. *Nature* **341**, 152–154 (1989).
- Sikdar, S. K., McIntosh, R. P. & Mason, W. T. *Brain Res.* **496**, 113–123 (1989).
- Bialojan, C. & Takai, A. *Biochem. J.* **256**, 283–290 (1988).
- Cohen, P. A. *Rev. Biochem.* **58**, 453–508 (1989).
- Armstrong, D. & Eckert, R. *Proc. natn. Acad. Sci. U.S.A.* **84**, 2518–2522 (1987).
- Armstrong, D. L. *Trends Neurosci.* **12**, 117–122 (1989).
- Reinhart, P. H., Chung, S., Martin, B. L., Brautigan, D. L. & Levitan, I. B. *J. Neurosci.* (in the press).
- Liebow, C., Reilly, C., Serrano, M. & Schally, A. V. *Proc. natn. Acad. Sci. U.S.A.* **86**, 2003–2007 (1989).
- Tashjian Jr. A. H. *Meth. Enzym.* **58**, 527–535 (1979).

ACKNOWLEDGEMENTS. We thank A. Lee for maintaining the GH_4C_1 cell line, R. Hall for engineering skills, and C. Miller and M. Garcia for generous gifts of charybdotoxin. A.S. was supported by the NIDDKD of NIH.

Sympathetic regulation of cardiac calcium current is due exclusively to cAMP-dependent phosphorylation

H. Criss Hartzell, Pierre-François Méry*, Rodolphe Fischmeister* & Gabor Szabo†

Department of Anatomy and Cell Biology, Emory University School of Medicine, Atlanta, Georgia 30322, USA

*Laboratoire de Physiologie Cellulaire Cardiaque, INSERM U-241, Université de Paris-Sud, F-91405 Orsay, France

†Department of Physiology, University of Virginia, Charlottesville, Virginia 22908, USA

THE positive inotropic effect of the sympathetic nervous system on the heart is partly mediated by an increase in the voltage-gated Ca^{2+} current (I_{Ca}). This increase is generally attributed to β -adrenergic receptor-stimulated cyclic AMP-dependent phosphorylation of the Ca^{2+} channel^{1,2}. It has been suggested that cAMP-dependent phosphorylation cannot explain all the effects of β -adrenergic agonists on I_{Ca} and that a parallel membrane-delimited pathway involving the 'direct' action of the G protein G_s also stimulates I_{Ca} (refs 3–7). A precedent exists for such a membrane-delimited pathway in the activation of a K^+ channel by acetylcholine in heart^{1,8}. A membrane-delimited pathway for stimulation of I_{Ca} might be important in rapid beat-to-beat regulation of contraction by the sympathetic nervous system⁹, because isoproterenol may produce a biphasic increase in I_{Ca} with the rapid phase ($\tau = 150$ ms) putatively mediated by the direct pathway and the slow phase ($\tau = 35$ s) by cAMP-dependent phosphorylation. Here we report that in frog, rat, and guinea pig ventricular myocytes I_{Ca} increases slowly and monophasically in response to isoproterenol. The increase is completely blocked by inhibitors of cAMP-dependent phosphorylation. Furthermore, the time course of the increase in I_{Ca} closely parallels the increase in contractile

Received 11 February; accepted 23 April 1991.

- Patel, Y. & Tannenbaum, G. S. (eds) *Metabolism* **39**, Suppl. 2, 1–191 (1990).
- Dorflinger, L. J. & Schonbrunn, A. *Endocrinology* **113**, 1541–1550 (1983).
- Dorflinger, L. J. & Schonbrunn, A. *Endocrinology* **113**, 1551–1558 (1983).

force produced by sympathetic nerve stimulation. These data refute earlier suggestions that regulation of Ca^{2+} channels by the sympathetic nervous system involves or requires a direct G-protein pathway.

Extremely rapid extracellular solution changes were produced by positioning a patch-clamped cell directly in front of one of several superfusion jets which could be changed in about 10 ms. The speed of solution change was measured as the change in holding current on switching between solutions with different potassium ion concentrations $[\text{K}^+]$ (Fig. 1a). The change in current began 5–10 ms after initiating the switch and was complete in 30–40 ms. Changing from one jet to another containing the same solution had no effect (for example, Fig. 1e, triangles).

Because the acetylcholine (ACh)-activated K^+ channel

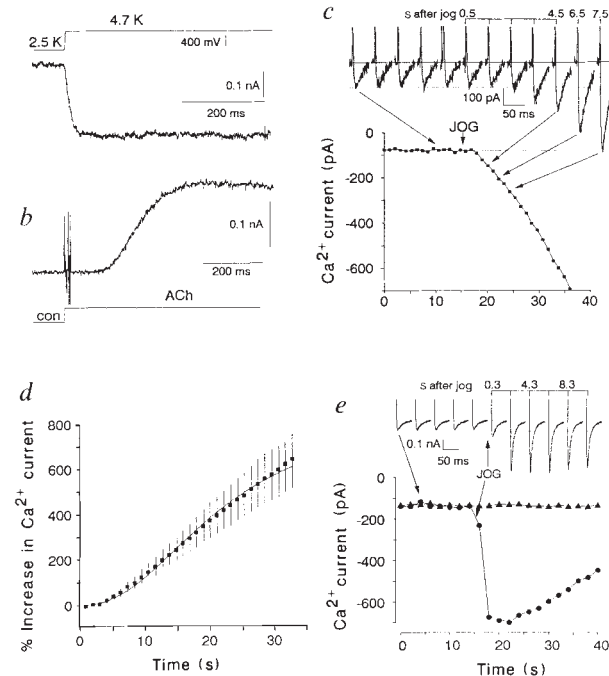


FIG. 1 Effects of rapid solution changes on ionic currents in frog cardiomyocytes. *a*, Changing external $[\text{K}^+]$. A frog ventricular myocyte was positioned at the end of superfusion jet containing 2.5 mM potassium Ringer's solution. The jets could be moved so that the cell was positioned in front of another jet with 4.7 mM potassium Ringer's. Top trace, output of the feedback circuit controlling jet position; bottom trace, current measured by patch-clamp. Holding potential, -100 mV. *b*, Rapid exposure of frog atrial myocyte to ACh. Cell was moved from control Ringer's to $1 \mu\text{M}$ ACh. *c*, Effect on I_{Ca} of a rapid switch to ISO in a frog ventricular myocyte. One I_{Ca} was elicited per s by 50-ms pulses from -80 mV to 0 mV. I_{Ca} was peak inward current and not leak-corrected. Current traces correspond to peak currents indicated on the graph. Switch from control Ringer's to $1 \mu\text{M}$ ISO occurred 450 ms after the fifth trace. Upper line, zero current; lower line, mean I_{Ca} before the rapid switch to ISO. *d*, Average from 10 experiments similar to that in *c*. Switch occurred at zero time. Error bars are s.e. *e*, Time course of effect of $(-)$ BAY K 8644 on I_{Ca} . I_{Ca} was elicited 0.5 s^{-1} by 200-ms pulses from -100 mV to 0 mV. Before the sixth trace (300 ms), the jets were moved to expose the cell either to $1 \mu\text{M}$ $(-)$ BAY K 8644 ($\bullet$) or Ringer ($\blacktriangle$) for 550 ms before returning to the first jet.

METHODS. Frog cardiomyocytes were isolated and whole-cell patch-clamped as described¹⁸. Solutions containing K^+ were used for *a*, *b*; all others used solutions containing Cs^+ . Internal Cs^+ (or K^+) solution (mM): 118 CsCl (or KCl), 4 MgCl_2 , 2.8 mM $\text{Na}_2\text{-K}_2\text{-ATP}$, 10 $\text{K}_2\text{-PIPES}$, 5 sodium creatine phosphate, 5 $\text{K}_2\text{-EGTA}$, pH 7.15 with KOH. Identical results were obtained with internal solutions containing 50–420 μM GTP and with $\text{pCa}^{2+}=10$ or 8.5. Potassium Ringer's (in mM), 123 NaCl, 2.5 (or 4.7) KCl, 10 HEPES, 1.8 CaCl_2 , 1.8 MgCl_2 , 5 pyruvic acid, 5 D-glucose , pH 7.4 with NaOH. Caesium Ringer's (in mM), 115 NaCl, 20 CsCl, 10 HEPES, 1.8 CaCl_2 , 1.8 MgCl_2 , 5 pyruvic acid, 5 D-glucose , pH 7.4 with NaOH. Cells were positioned at the end of one of several capillaries (400 μm inside diameter, flow rate 5 cm s^{-1}) mounted on a stepper motor¹⁹ or a loudspeaker coil²⁰ that could move between capillaries in <20 ms and in <5 ms, respectively.

($I_{\text{K(ACh)}}$) is activated directly by a G protein^{1,8} this activation rate can indicate how rapidly a direct G-protein pathway could act (Fig. 1b). $I_{\text{K(ACh)}}$ began to activate about 100 ms after the solution change and was maximally activated in 400 ms. The kinetics are similar to those obtained with iontophoretic application of ACh (ref. 10).

The rate of $I_{\text{K(ACh)}}$ activation was at least 20 times more rapid than the effect of isoproterenol (ISO) on I_{Ca} . In the experiment shown in Fig. 1c, I_{Ca} was elicited per second in frog myocytes by 50-ms pulses from -80 mV to 0 mV. After switching to a solution containing $1 \mu\text{M}$ ISO, I_{Ca} was unchanged for the first three pulses. Thereafter, I_{Ca} increased monophasically. ISO caused a maximal 8.5-fold increase in I_{Ca} , but a significant ($P < 0.05$) increase (6.7%) was not observed until 3 s after switching to ISO ($n = 10$; Fig. 1d). Similar results were obtained when Ba^{2+} replaced Ca^{2+} as charge carrier. The increase in I_{Ca} produced by forskolin was about twice as slow as the response to ISO (not shown), presumably because forskolin binds slowly to adenylyl cyclase.

The slowness of the response of I_{Ca} to ISO was not due to an inherent slowness in the responsiveness of Ca^{2+} channels to agonists, because I_{Ca} increased very rapidly in response to the dihydropyridine $(-)$ BAY K 8644 (Fig. 1e). After switching to $1 \mu\text{M}$ $(-)$ BAY K 8644, I_{Ca} almost doubled in amplitude by 300 ms and reached a maximum of about 5-fold increase by 2.3 s. By contrast, at 3 s ISO stimulated I_{Ca} less than 10% (Fig. 1d). The response to ISO had a similar time course when the cells were perfused internally with a submaximal concentration of cAMP to phosphorylate Ca^{2+} channels. Thus this excludes the possibility that the direct G-protein pathway requires phosphorylation.

The slowness of the response of I_{Ca} to ISO relative to the rapid response of $I_{\text{K(ACh)}}$ to ACh suggests either that there is no direct G-protein regulation of I_{Ca} in frog or that the direct G-protein effect on I_{Ca} is much slower than activation of $I_{\text{K(ACh)}}$. If there were direct G-protein effects on I_{Ca} , one would predict that inhibitors of cAMP-dependent protein kinase (A-PK) would not completely block the effects of ISO. Internal perfusion with adenosine 3',5'-(Rp)-phosphothioate (Rp-cAMP-S), a competitive inhibitor of A-PK (ref. 11), decreased I_{Ca} elevated by ISO to control levels (Fig. 2a). On average, $1 \mu\text{M}$ ISO stimulated I_{Ca} in frog $10.8\text{-fold} \pm 2.2\text{-fold}$ and 1 mM Rp-cAMP-S blocked the stimulation $97.5\% \pm 3.6\%$ ($n = 5$). Similar results were obtained with the A-PK peptide inhibitor PKI(5-22) (ref. 12). PKI(5-22) ($16 \mu\text{M}$) blocked $100\% \pm 1\%$ ($n = 5$) of the ISO-stimulated current. Although it is clear that steady-state ISO-stimulated current is blocked by PKI, a direct G-protein regulation of I_{Ca} might be transient and only be seen at short times after ISO application. To test this possibility, 6-s applications of ISO were delivered at various times during internal perfusion with PKI(15-22) (ref. 12) (Fig. 2b). Within 17 min of PKI(15-22) perfusion, the response to ISO was completely blocked. This result excludes a transient, direct G-protein effect.

These results show clearly that there is no direct G-protein modulation of I_{Ca} in frog ventricular myocytes. But because frog heart beats spontaneously at a frequency of $<1 \text{ s}^{-1}$, rapid G-protein regulation of I_{Ca} may be less pronounced than in a species with a higher beat frequency. But no evidence was found for a rapid, direct G-protein effect in ventricular myocytes isolated from rat heart, which beats at $>5 \text{ s}^{-1}$ (Fig. 2c). In this cell, I_{Ca} began to increase only about 4 s after the switch to ISO. The increase was monophasic. As observed in frog, the increase in I_{Ca} produced by ISO was completely blocked by internal perfusion with Rp-cAMP-S (Fig. 2d). Assuming that I_{Ca} run-down occurred at a constant rate during an experiment, $1 \mu\text{M}$ ISO increased I_{Ca} in rat $127\% \pm 13\%$ and 5 mM Rp-cAMP-S decreased the ISO-stimulated current $117\% \pm 9\%$ (slightly below the control level, $n = 7$). Thus, ISO stimulation of I_{Ca} in rat, as well as frog, can be explained entirely by cAMP-dependent phosphorylation. We also performed experiments with guinea

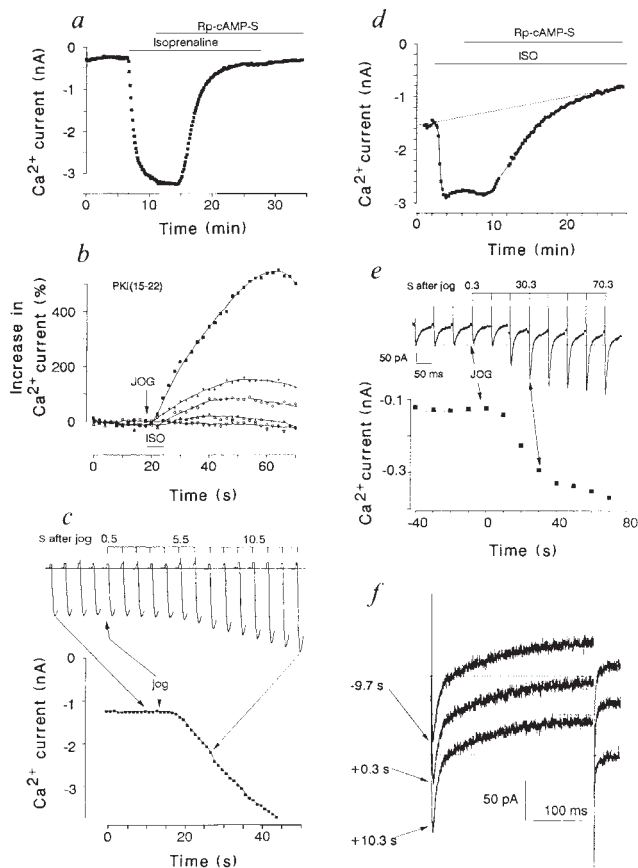


FIG. 2 Effects of inhibitors of cAMP-dependent protein kinase. *a*, Effects of Rp-cAMP-S on ISO-elevated I_{Ca} in frog ventricular myocyte. I_{Ca} was evoked 1 per 8 s by 400-ms pulses from -80 mV to 0 mV. I_{Ca} was net inward current. ISO ($1 \mu\text{M}$) and Rp-cAMP-S (1 mM) were present as indicated. *b*, Effects of peptide A-PK inhibitor on response of I_{Ca} to a blast of ISO in a frog ventricular myocyte. Blasts (6-s duration) of ISO (horizontal bar) were given at different times after the establishment of whole cell recording with $16 \mu\text{M}$ PKI (sequence, RTGRRNAL; single-letter amino-acid notation) in the patch pipette. The switch to ISO occurred at $t = 18.4$ s. Time after patchbreak: $\blacksquare$, 2.5 min; $+$, 6 min; $\diamond$, 10 min; $\triangle$, 13 min; ∇ , 17 min. *c*, Effect of ISO blast on a rat ventricular myocyte. I_{Na} was blocked by $30 \mu\text{M}$ tetrodotoxin (TTX) and zero-sodium Ringer. I_{Ca} was evoked per s by 20-ms pulses from -90 mV to -15 mV. Traces correspond to the peak inward currents indicated in the graph. The step to $1 \mu\text{M}$ ISO occurred 500 ms after the 4th trace shown. *d*, Effect of Rp-cAMP-S on ISO stimulation of I_{Ca} in rat ventricular myocyte. I_{Ca} was elicited 0.1 s^{-1} by a 200-ms pulse from -80 mV to 0 mV. External solution was caesium Ringer's (see Fig. 1). I_{Na} was blocked by $10 \mu\text{M}$ TTX and a 10-ms prepulse to -50 mV. I_{Ca} rundown was extrapolated from rundown before ISO exposure (dotted line). ISO ($1 \mu\text{M}$) and Rp-cAMP-S (5 mM) were present as indicated. *e*, Effect of rapid exposure of guinea pig ventricular myocyte to $2 \mu\text{M}$ ISO. Cell was stimulated 0.1 s^{-1} by a 300-ms pulse from -50 mV holding potential to 0 mV. *f*, Currents recorded 9.7 s before and 0.3 s and 10.3 s after the switch to ISO are superimposed with a small offset for clarity. The increase in holding current that develops ~ 20 s after the pulse is probably the ISO-activated Cl^- current.

METHODS. Cells were perfused internally as described²¹. Rat and guinea pig ventricular cells were isolated by modification of the methods of Mitra and Morad²². Rat internal solution (mM): 110 CsCl, 4 Cs₂-K₂-ATP, 10 K₂-EGTA, 1 caesium 1,2-Bis(*O*-aminophenoxy)ethane-*N,N,N',N'*-tetracetic acid (Cs₄-BAPTA), 6 MgCl₂, 6.25 Tris-creatine phosphate, pH 7.15 with CsOH. Zero sodium Ringer's, 120 tetraethylammonium (TEA) chloride, 1.8 CaCl₂, 1.8 mM MgCl₂, 10 HEPES, 5 pyruvic acid, 5 D-glucose, pH 7.4 with TEA-OH. Guinea pig external solution: 127.1 NaCl, 20 CsCl, 4 NaHCO₃, 0.8 NaH₂PO₄, 1.8 MgCl₂, 1.8 CaCl₂, 5 D-glucose, 5 sodium pyruvate, 5×10^{-2} TTX, 10 HEPES pH 7.4 with NaOH. Guinea pig internal solution, 139.8 CsCl, 5 potassium EGTA, 4 MgCl₂, 5 sodium creatine phosphate, 3.1 sodium ATP, 0.42 sodium GTP, 0.062 CaCl₂ ($p\text{Ca} = 8.5$), 10 HEPES pH 7.1 with KOH.

pig myocytes, because studies showing membrane-delimited regulation of cardiac Ca^{2+} channels were performed in this species^{6,7,9}. The stimulation frequency and timing of the ISO application were the same as used by Yatani and Brown⁹; I_{Ca} was evoked every 10 s and the cell was rapidly switched to the ISO jet 300 ms before a stimulus. Contrary to the findings of Yatani and Brown⁹, I_{Ca} did not increase until the second (10.3 s) or third (20.3 s) pulse after the rapid switch to ISO (Fig. 2*e*), and I_{Ca} evoked 300 ms and 10.3 s after ISO application had the same time course as control I_{Ca} (six ventricular and three atrial cells) (Fig. 2*f*).

To test whether direct G-protein effects might be necessary to ensure rapid responses of heart to sympathetic nerve stimulation⁹, the rate of change in contractile force of isolated frog hearts upon nerve stimulation was determined. Responses to parasympathetic stimulation depended on the timing of the stimulus to the beat cycle, but inhibitory effects on contractile activity could be recorded < 1 s after stimulation (Fig. 3*a*). By contrast, after sympathetic stimulation the force of contraction increased noticeably only after several beats (Fig. 3*b, c*). On average, increases in the force of contraction did not start until 7 s after sympathetic stimulation (Fig. 3*d*). Because activation of the contractile apparatus in frog heart occurs largely through influx of Ca^{2+} from the extracellular space, rather than release from internal stores, contraction should be a reliable index of I_{Ca} (ref. 13). Analogous results have been obtained by us in intact rat, and by others in frog¹⁴ and rabbit isolated sinoatrial node¹⁵, where effects of sympathetic stimulation occurred five times more slowly than those of parasympathetic stimulation.

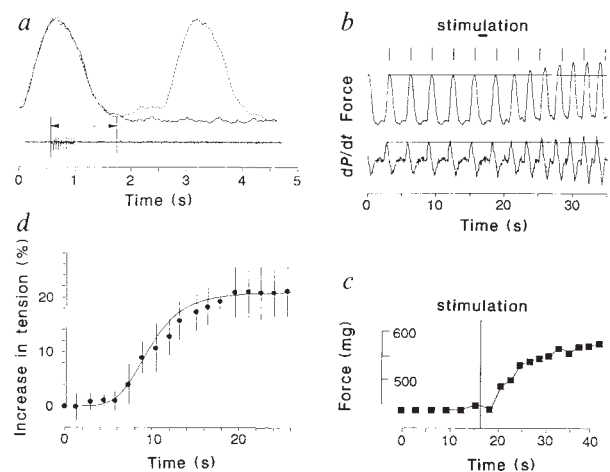


FIG. 3 Time course of change in contractile force of frog heart with nerve stimulation. *a*, Effect of parasympathetic nerve stimulation. Bottom trace, vagus nerve stimulation for 500 ms at 20 Hz. Upper traces, two superimposed traces: solid trace, tension record obtained concurrently with the stimulus trace; broken trace, two contractions before nerve stimulation to indicate when a contraction would have been expected to occur if the nerve had not been stimulated. The two traces diverge about 1 s after onset of nerve stimulation. *b*, Time course of the effect of sympathetic nerve stimulation on force of contraction. The heart was perfused with Ringer's solution containing $1 \mu\text{M}$ atropine and the vagosympathetic nerve was stimulated at 20 Hz as indicated. *c*, Peak force versus time for the experiment in *b*. *d*, Average time course of effect of sympathetic nerve stimulation on force. The graph shows the mean per cent increase in peak tension $\pm$ s.d., as a function of time after sympathetic nerve stimulation. Stimuli were applied randomly during the cardiac cycle and data from 14 stimuli (three hearts) were averaged into 1.5-s bins for plotting.

METHODS. Frog hearts were dissected with attached nerves for stimulation by a suction electrode, perfused in a Langendorff apparatus, and attached to a force transducer. Sympathetic and parasympathetic nerves could be stimulated separately by dissecting the vagosympathetic nerve to the medulla, where the vagus nerve could be separated from the ascending paravertebral sympathetic nerves. Similar results were obtained when sympathetic and parasympathetic responses were separated pharmacologically with propranolol and atropine.

Our results show clearly that the regulation of I_{Ca} by β -adrenergic receptors does not involve a direct G-protein pathway and that such a pathway is not required to explain the physiological regulation by the sympathetic nervous system. Studies showing that the effect of ISO is not additive with the effect of internally perfused cAMP of forskolin¹⁶⁻¹⁸ are consistent with this conclusion. It is unlikely that our recording conditions inhibited the direct G-protein effect because $I_{K(ACh)}$ activated normally, and the positive inotropic response in intact hearts and the response of I_{Ca} to ISO in isolated myocytes had similar latencies. Our results contradict those of Yatani and Brown⁹. But in these studies, it was not shown that the rapid increase in I_{Ca} (which was presumed to be mediated by the direct pathway) was not blocked by inhibitors of A-PK. Thus, the rapid increase could be explained by mechanisms other than a direct

G-protein pathway.

Although direct G-protein regulation of I_{Ca} seems to have no role in β -adrenergic regulation of I_{Ca} *in vivo*, the possibility remains that G proteins do stabilize or affect Ca^{2+} channel function. Evidence suggesting a direct G protein pathway for I_{Ca} regulation comes from cell-free experiments with artificial membranes or excised membrane patches²⁻⁵. Studies on intact cells have used very nonphysiological ionic and metabolic conditions (cells were depleted of ATP and exposed to Na^+ - and K^+ -free solutions)^{6,7}. Under these conditions, GTP- γ -S or pre-activated G_s protein can increase I_{Ca} by about 50%. Under more physiological conditions, the Ca^{2+} channel might already be associated with G protein such that cAMP-dependent phosphorylation is the only regulatory mechanism activated by β -adrenergic receptors. □

Received 8 April; accepted 21 May 1991.

- Hartzell, H. C. *Prog. Biophys. molec. Biol.* **52**, 165-247 (1988).
- Trautwein, W. & Hescheler, J. A. *Rev. Physiol.* **52**, 257-274 (1990).
- Yatani, A. *et al. Science* **238**, 1288-1292 (1987).
- Yatani, A. *et al. J. Biol. Chem.* **263**, 9887-9895 (1988).
- Imoto, Y. *et al. Am. J. Physiol.* **255**, H722-H728 (1988).
- Shuba, Y. M., Hesslinger, B., Trautwein, W., McDonald, T. & Pelzer, D. *J. Physiol., Lond.* **424**, 205-228 (1990).
- Pelzer, S. *et al. Am. J. Physiol.* **259**, H264-H267 (1990).
- Breitwieser, G. E. & Szabo, G. *Nature* **317**, 538-540 (1985).
- Yatani, Y. & Brown, A. M. *Science* **245**, 71-74 (1989).
- Simmons, M. A. & Hartzell, H. C. *Pflügers Arch.* **409**, 454-461 (1987).
- Botelho, L. H. P., Rothermel, J. D., Coombs, R. V. & Jasorff, B. *Meth. Enzym.* **159**, 159-172 (1988).
- Kemp, B. E., Cheng, H. C. & Walsh, D. A. *Meth. Enzym.* **159**, 173-182 (1988).

- Morad, M., Goldman, Y. E. & Trentham, D. R. *Nature* **304**, 635-638 (1983).
- Hutter, O. & Trautwein, W. *J. gen. Physiol.* **39**, 715-733 (1956).
- Spear, J. F., Kronhaus, K. D., Moore, E. N. & Kline, R. P. *Circulation Res.* **44**, 75-88 (1979).
- Kameyama, M., Hofmann, F. & Trautwein, W. *Pflügers Arch.* **407**, 123-128 (1986).
- Hartzell, H. C. & Fischmeister, R. *Molec. Pharmacol.* **32**, 639-645 (1987).
- Fischmeister, R. & Hartzell, H. C. *J. Physiol., Lond.* **376**, 183-202 (1986).
- Hartzell, H. C. & Budnitz, D. *Biophys. J.* **59**, 551 (1991).
- Méry, P.-F., Lechêne, P., Levi, R. C. & Fischmeister, R. *Biophys. J.* **59**, 551 (1991).
- Hartzell, H. C. & Fischmeister, R. *Nature* **323**, 273-275 (1986).
- Mitra, R. & Morad, M. *Am. J. Physiol.* **249**, H1056-H1060 (1985).

ACKNOWLEDGEMENTS. We thank I. Cohen for Rp-cAMP-S, D. Glass for PKI and W. Goolsby, D. Budnitz, P. Richer and P. Lechêne for assistance. Supported by the NIH (H.C.H. and G.S.), Emory University Research Committee (H.C.H.), Bayer-Pharma (R.F.) and MRT (France) (G.S.).

Differential regulation of rasGAP and neurofibromatosis gene product activities

Gideon Bollag & Frank McCormick

Department of Molecular Biology, Cetus Corporation, 1400 53rd Street, Emeryville, California 94608, USA

THE *ras*-encoded p21^{ras} proteins bind GTP very tightly, but catalyse hydrolysis to GDP very slowly¹. In humans, two genes encode proteins that stimulate this GTPase activity (GAP, or GTPase-activating proteins)², one of relative molecular mass 120,000, referred to as p120-GAP, and another NF1-GAP, which is encoded by the neurofibromatosis type-1 gene³⁻⁵. Both GAPs are widely expressed in mammalian tissues^{6,7}. Here we show that although they will both bind oncogenic mutants of p21^{ras}, neither will stimulate their GTPase activity. NF1-GAP binds to the p21^{ras} proteins up to 300 times more efficiently than p120-GAP. The two GAPs are inhibited to different extents by certain lipids: micromolar concentrations of arachidonate, phosphatidate and phosphatidylinositol-4,5-bisphosphate affect only NF1-GAP. This inhibition does not compete with p21^{ras}, and lipid-inactivated NF1-GAP can still bind p21^{ras}. We used the detergent dodecyl maltoside, which inhibits only NF1-GAP, to distinguish between the two activities in cell extracts and found both types present together in several mammalian cell lines. In contrast, GAP activity in extracts of *Xenopus* oocytes was not affected by dodecyl maltoside. By these criteria, the mammalian cells contain both GAP activities and the oocytes have only p120-like GAP activity. These results indicate that more than one GAP regulates p21^{ras} in the same cell.

As it is the GTP-bound form of p21^{ras} that is active in signal transduction⁸, one function of GAP is to terminate signalling. In addition, several lines of evidence suggest that p120-GAP acts downstream of p21^{ras} (ref. 2). First, p120-GAP shows greatly reduced activity on biologically inactive p21^{ras} proteins with mutations in a region that is presumed to interact with a downstream effector^{9,10}. Second, although p120-GAP does not catalyse the hydrolysis of GTP on oncogenic p21^{ras} proteins⁸,

it still binds to these proteins¹¹. Although this binding does not effect downregulation, the p21^{ras}/GAP complex may still have a role downstream. Third, GAP is necessary for p21^{ras} to inhibit G-protein activation of atrial potassium channels¹². Preliminary work has indicated that effector mutants are also poor substrates for NF1-GAP (refs 3, 4). We now investigate whether NF1-GAP can also function as signal transducer.

Table 1 lists a comparison of GAP activities on wild-type, oncogenic and effector mutants of p21^{ras}. The intrinsic GTPase activities, the binding affinities to p120- and NF1-GAP, and the stimulated GTPase activities are shown. We have previously demonstrated that the catalytic region of NF1-GAP (the GAP-related domain, referred to as NF1-GRD) binds better than p120-GAP to wild-type N-ras, and that both GAPs are ineffective in stimulating the GTPase activities of mutant p21^{ras} proteins³. Here we compare GAP binding to wild-type and mutant p21^{ras} proteins. The wild-type proteins were N-ras and H-ras, two of the human *ras* proto-oncogene products. We found that p120-GAP bound to both proteins with comparable affinity, but that NF1-GAP had a fourfold higher affinity for H-ras. As the regions of p21^{ras} thought to interact with p120-GAP are identical in N-ras and H-ras (ref. 13), there must be other regions involved as well. Like the wild-type proteins, the oncogenic p21^{ras} proteins (N-ras with Gly 12 replaced by Asp, and H-ras with Gln 61 replaced by Leu or His) bound to NF1-GAP with much greater affinity than to p120-GAP, but the GTPase-stimulating activity of both GAPs was reduced by >1,000-fold. One biologically inactive protein (effector mutant, Asp 38 replaced by Ala) had reduced affinity towards both GAPs. Although the affinity of this mutant for NF1-GAP was reduced by over 100-fold, it still bound about as tightly to NF1-GAP as wild-type protein did to p120-GAP. To interpret this in terms of a possible signalling role for NF1-GAP, NF1-GAP may require a higher occupancy of p21^{ras} than p120-GAP, affinities might be regulated by other factors in the cell, or the affinity need not directly determine activation. The results are anyway consistent with the hypothesis that both GAPs function downstream of p21^{ras}: they both bind well to constitutively signalling oncogenic p21^{ras} proteins but interact poorly with the effector mutant in which signalling is impaired.

Although there is no direct evidence for a role of GAP in

TABLE 1 Properties of p21^{ras} proteins and their interactions with p120- and NF1-GAPs

p21 ^{ras} protein*	Intrinsic GTPase [†] (mmol min ⁻¹ mol ⁻¹)	p120-GAP binding [‡] (μM)	NF1-GAP binding [‡] (μM)	p120-GAP activity§ (% relative to N-ras)	NF1-GAP activity§ (% relative to N-ras)
N-ras (wild-type)	3.55	5	0.16	100	100
N-ras G12D	1.50	5	0.16	<0.0001	0.018
H-ras (wild-type)	3.92	5	0.035	96	96
H-ras D38A	5.35	>10	5	0.53	8.1
H-ras Q61L	0.056	0.5	0.002	0.0087	<0.0001
H-ras Q61H	0.36	2	0.025	0.027	0.072

* GTPase activities and GAP binding values were determined for these p21^{ras} proteins. N-ras proteins and H-ras D38A were purified by anion-exchange and size-exclusion chromatography. The other H-ras proteins were purified from insoluble fractions of *Escherichia coli* lysates by extraction with 20 mM NaOH followed by neutralization with 50 mM Tris, pH 7.5, 5 mM MgCl₂, 2 mM dithiothreitol (DTT), 0.1 mM GDP. All proteins were >90% pure after anion-exchange chromatography.

[†] Intrinsic GTPase activities were determined by incubating 5 nM p21^{ras}. GTP ([γ-³²P]GTP; 6,000 Ci mmol⁻¹) in 20 mM HEPES, pH 7.5, 2 mM MgCl₂, 2 mM DTT at 22 °C for 90 min. Phosphate release²³ in all cases was less than 25% of total GTP.

[‡] Binding affinity was determined from competitive inhibition by mutant p21^{ras}. GTP-γ-S versus wild-type N-ras.GTP ([γ-³²P]GTP; 6,000 Ci mmol⁻¹) (ref. 11). Different concentrations of p21^{ras} proteins prebound to GTP-γ-S were added to 1.25 nM wild-type N-ras.GTP, 20 mM HEPES, pH 7.5, 2 mM MgCl₂, 2 mM DTT and 0.25 nM p120-GAP or 1 nM NF1-GRD and incubated for 5 min at 22 °C. Binding is expressed as the micromolar p21^{ras}.GTP-γ-S protein concentration necessary for half-maximal inhibition. Levels had to be much higher for competitive inhibition by the GDP-bound form of the proteins; for example, the affinity of Q61L.GDP for NF1-GRD was ~250 nM.

§ GAP activities are normalized to p120-GAP or NF1-GRD activity on wild-type N-ras protein. The actual values under these conditions (13 fmol min⁻¹ fmol⁻¹ p120-GAP, and 6 fmol min⁻¹ fmol⁻¹ NF1-GRD) are very dependent on GAP concentration. 50 nM p21^{ras}.GTP was incubated with 0.2–100 nM GAP (optimized to yield <25% hydrolysis for the particular mutant) for 5 min at 22 °C in 20 mM HEPES, pH 7.5, 1 mM MgCl₂, 2 mM DTT.

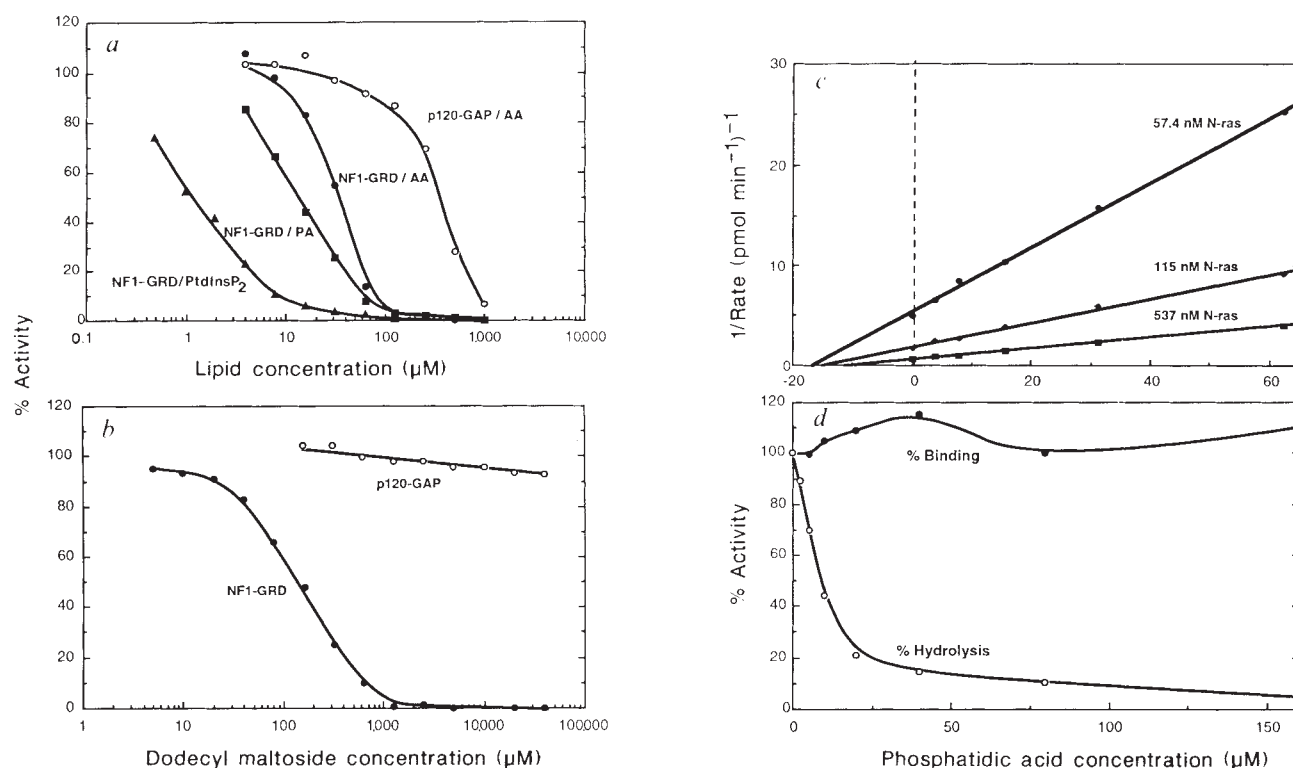
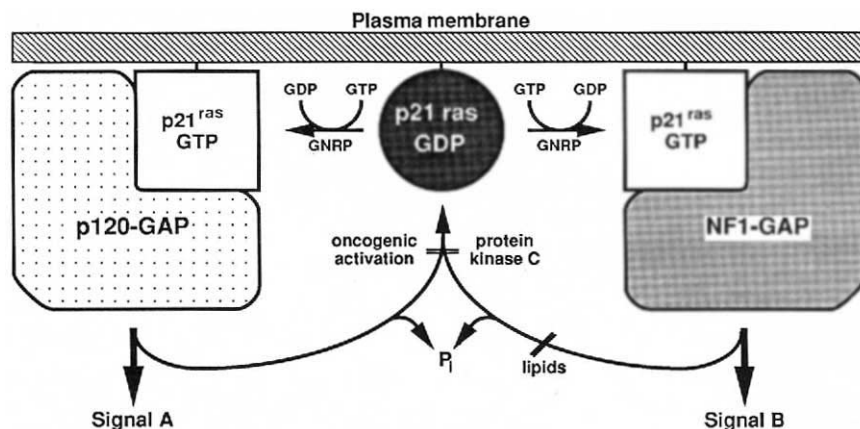


FIG. 1 *a* and *b*, Selective inhibition of NF1-GAP activity. *a*, Inhibition of p120-GAP by arachidonic acid (p120-GAP/AA) and inhibition of NF1-GRD by arachidonic acid (NF1-GRD/AA), phosphatidyl inositol 4,5-bisphosphate (NF1-GRD/PtdInsP₂), and 1-stearoyl-2-arachidonoyl phosphatidic acid (NF1-GRD/PA) are labelled on the graph. *b*, Inhibition of NF1-GRD but not p120-GAP by *n*-dodecyl β-D-maltoside. This inhibition uses only the catalytic region of the NF1 gene; we observe that GAP activity in crude extracts can be inhibited by comparable levels of lipid or detergent, suggesting that intact NF1-GAP has a similar sensitivity to lipids. *c* and *d*, Phosphatidic acid and p21^{ras} bind noncompetitively. *c*, Dixon plot indicating noncompetitive inhibition of NF1-GRD by 1-stearoyl-2-arachidonoyl phosphatidic acid. The lines intersect at the x axis, indicating a constant *K*_i of ~12 μM. *d*, Per cent GTP hydrolysis on H-ras Q61H catalysed by NF1-GRD (% hydrolysis), or binding of H-ras Q61H to NF1-GRD (% binding) is shown as a function of phosphatidic acid concentration.

METHODS. Inhibition of GAP activity by selected lipids or detergent was determined at the indicated concentrations in 0.1% NP-40, 20 mM Tris, pH 7.5, 2 mM MgCl₂, 2 mM DTT. Concentrations of N-ras.GTP ranged from 10 to 30 nM, and GAP concentrations from 0.5 to 1 nM. GAP activity was determined by assaying for phosphate released from [γ-³²P]GTP

(6,000 Ci mmol⁻¹) during a 10-min reaction at 22–25 °C (ref. 23). Results were the same whether or not lipid solutions were sonicated. Data for lipids that did not significantly inhibit p120-GAP (phosphatidyl inositol 4,5-bisphosphate, 1-stearoyl-2-arachidonoyl phosphatidic acid, 1-stearoyl-2-arachidonoyl phosphatidyl choline, 1-stearoyl-2-arachidonoyl diacylglycerol, and arachidonic acid) or NF1-GRD (1-stearoyl-2-arachidonoyl phosphatidyl choline, 1-stearoyl-2-arachidonoyl diacylglycerol, and arachidonic acid) are not shown. *c* and *d*, GAP activity determined as in *a* and *b*, except that different concentrations of N-ras and GTP were used. Concentrations of N-ras.GTP are labelled in *c*. [γ-³²P]GTP specific activities ranged from 60–270 Ci mmol⁻¹. For the binding assay in *d*, NF1-GRD containing an epitope tag was immobilized on sepharose beads covalently bound to the antibody against this epitope³. These beads (~2 nM NF1-GRD) were incubated with H-ras Q61H bound to [γ-³²P]GTP (38 nM) at the indicated concentrations of phosphatidic acid in 0.1% NP-40, 20 mM HEPES, pH 7.5, 2 mM MgCl₂, 2 mM DTT. After 5 min at 23 °C, beads were washed three times and counted for ³²P. Background was less than 2% as determined in an identical reaction containing 2% SDS. The per cent hydrolysis during 30 min at 23 °C using 25 nM H-ras Q61H and 50 nM NF1-GRD was determined in the same buffer.

FIG. 2 Model depicting possible roles and regulation of p120-GAP and NF1-GAP. p21^{ras}.GDP is converted to p21^{ras}.GTP through the action of a guanine nucleotide-releasing protein (GNRP). Normally, both p120-GAP and NF1-GAP collaborate to stimulate GTP hydrolysis, releasing inorganic phosphate (P_i) and returning p21^{ras} to the GDP-bound state. p21^{ras}.GTP binds to both GAPs, and the resulting complexes would be competent to generate putative signals A and B. Alternatively, the signals could be generated independently of GAP. As described in the text, oncogenic activation of p21^{ras} and protein kinase C may inhibit either GTPase-stimulating activity, whereas lipids primarily inhibit NF1-GAP.



signal transduction, indirect evidence has come from studying the regulation of GAP activity. For example, lipids that perform second messenger signalling roles inhibit GAP activity in cell extracts¹⁴⁻¹⁶. Furthermore, protein kinase C activation leads to the inhibition of GAP activity in T lymphocytes¹⁷. Indeed, it seems that p21^{ras} is regulated primarily by control of GAP activity in these cells. We investigated which of the GAP activities were affected by the lipids or by protein kinase C.

As shown in Fig. 1a, the lipid inhibition of GAP activity affects NF1-GAP primarily. Phosphatidate, arachidonate, and phosphatidylinositol 4,5-bisphosphate inhibit NF1-GRD, whereas phosphatidylcholine, arachidate, and diacylglycerol do not; p120-GAP activity is only weakly affected by these lipids. Phosphatidylinositol-4,5-bisphosphate has the highest affinity for NF1-GRD (concentration of 50% inhibition, $K_i \approx 1 \mu\text{M}$). Interestingly, the actin-binding protein profilin, which also binds tightly to phosphatidylinositol-4,5-bisphosphate, inhibits phospholipase C activity by this interaction¹⁸. The K_i for inhibition by phosphatidate was higher ($\sim 12 \mu\text{M}$). But as p21^{ras} is located at the cytoplasmic membrane, local concentrations of fluxing lipids might reach these levels.

To clarify the significance of the lipid inhibition, we investigated whether p21^{ras} could still bind to lipid-inactivated NF1-GAP. Half-maximal inhibition by phosphatidic acid was measured over a range of N-ras concentrations. A plot of reciprocal activity versus inhibitor concentration suggested inhibition was noncompetitive (Fig. 1c). To test the prediction that there are distinct binding sites for lipids and p21^{ras} on NF1-GAP, a direct binding assay was developed which indicated that phosphatidic acid did not block binding of H-ras Q61H to NF1-GAP at concentrations that completely blocked NF1-GAP activity on this same mutant (Fig. 1d). This mutant was used as it bound tightly to NF1-GAP and gave low but detectable NF1-GAP-stimulated GTP hydrolysis. Thus in the presence of these lipids, NF1-GAP can associate with p21^{ras} without inducing hydrolysis of GTP. By trapping p21^{ras} in the GTP-bound form, these lipids can function in the cell to activate the p21^{ras}/GAP signalling pathways.

Because inhibition by lipids discriminates between p120-GAP and NF1-GAP activities, it is possible to determine which activity resides in which cell type. Owing to the instability of unsaturated lipids, we used a more stable, strikingly selective inhibitor of NF1-GAP, namely the detergent *n*-dodecyl β -D-maltoside (Fig. 1b). The K_i for inhibition of NF1-GAP was $\sim 200 \mu\text{M}$, but p120-GAP remained active even at 40 mM dodecyl maltoside. We therefore used this detergent to distinguish between p120-GAP and NF1-GAP activities in crude cell extracts of established cell lines.

As shown in Table 2, all of the mammalian cells contained inhibitable (NF1-like) and uninhibitable (p120-like) GAP activities in varying proportions. Cells tested included: NIH 3T3 cells and a v-src-transformed derivative (SRD 3T3); PC 12 cells

(which differentiate in response to activated p21^{ras}); two human cell lines containing activated N-ras alleles (PA 1: teratocarcinoma with G12D N-ras; HL 60: acute myeloblastic leukaemia with Q16L N-ras); Jurkat cells (derived from a human T-cell leukaemia); and peripheral blood lymphoblasts (PBLs). Previous work has shown that phorbol esters cause an inhibition of GAP activity in Jurkat cells and PBLs (ref. 17). Here we show that both p120- and NF1-like GAP activities are affected by this treatment, but the inhibition is insufficient to account for the large difference in GTP-bound p21^{ras} in phorbol ester-treated cells¹⁷. This indicates that the inhibition could be caused by a diffusible factor (which would be highly diluted in our extracts), rather than by direct phosphorylation of the GAPs. Also shown in Table 2 are results from extracts of bovine brain and human placenta. Interestingly, human placenta contains large amounts of p120-like GAP and small amounts of NF1-like GAP. When exogenous NF1-GRD was added to some of the cell extracts, the increased activity could still be inhibited by dodecyl maltoside, suggesting that differences in detergent sensitivity were not due to differences in amounts of bulk lipids. In summary, even clonal cell lines seem to have more than one type of GAP activity.

In *Xenopus laevis* oocytes p21^{ras} induces meiosis¹⁹ and as these cells contain GAP activity⁸, we investigated this activity further. Also shown in Table 2 are data indicating that GAP activity in extracts of stage-VI *Xenopus* oocytes is not inhibited by dodecyl maltoside, suggesting that these cells contain only p120-like-GAP. If p120-GAP does indeed act downstream of p21^{ras}, then it may be on the signalling pathway that leads to meiosis.

In addition to comparing GAP activities in these cell lines, we also looked at their membrane localization. NIH 3T3 cells and bovine brain extracts were fractionated and assayed as described in the legend to Table 2. No significant differences were noted in the distribution of NF1-like- and p120-like-GAP activities. Interestingly, IRA2, a yeast homologue of GAP (ref. 20), may be membrane-bound²¹. Although NF1-GAP shares more extensive homology with IRA2 than with p120-GAP (refs 5,7), NF1-GAP activity does not preferentially localize to the membrane.

The identification of a second rasGAP was unexpected and any explanation must account for the fact that both activities can reside in a single cell. These GAPs evidently possess signalling functions in addition to their ability to downregulate p21^{ras}. Results from p120-GAP overexpression argue against the hypothesis that p120-GAP is the only downstream target of p21^{ras} (ref. 22). Indeed, the pleiotropic effects of p21^{ras} activation suggest that there are several downstream targets. As mammalian cells have at least two distinct GAP activities, it is possible that the multiple effectors are different GAPs; such a model is depicted in Fig. 3. This model predicts that oncogenic activation would turn on all GAP signals by trapping p21^{ras} in its GTP-bound form; as shown here, the oncogenic proteins still bind

TABLE 2 p120-GAP-like and NF1-GAP-like activities in different cell extracts

Extract	Total activity (relative to p120-GAP)	Non-inhibitable p120-GAP-like (relative to p120-GAP)	Inhibitable NF1-GAP-like (relative to p120-GAP)
1 nM p120-GAP	100	103.8	—
1 nM NF1-GRD	60.95	—	60.95
NIH 3T3 cells	61.77	25.38	36.39
SRD 3T3 cells	88.62	22.26	66.36
PC 12 cells	104.9	25.99	78.95
Jurkat cells	71.97	30.62	41.35
PMA-treated Jurkat cells	44.18	24.45	19.73
PBLs	116.0	66.17	49.87
PMA-treated PBLs	46.50	16.98	29.52
PA 1 cells	150.7	101.2	49.45
HL 60 cells	65.65	36.50	29.15
Stage-VI oocytes	20.00	19.43	0.57
Brain cytosol	86.48	68.61	17.87
Brain membranes	52.69	41.15	11.54
Human placenta	393.1	354.6	38.49

Data are expressed as activity relative to 1 nM p120-GAP under the same conditions. Cultured cells were lysed in 2 to 4 volumes of 1% Nonidet P-40 (NP-40), 20 mM HEPES, pH 7.5, 20 mM EDTA, 2 mM DTT, 10 mM sodium fluoride, 1 mM sodium orthovanadate, 1 mM PMSF, 100 $\mu\text{g ml}^{-1}$ leupeptin, 100 $\mu\text{g ml}^{-1}$ pepstatin, 1 $\mu\text{g ml}^{-1}$ okadaic acid. Phorbol myristyl acetate (PMA)-treated cells were incubated with 100 ng ml^{-1} PMA for 30 min before collecting. One hundred stage-VI *Xenopus* oocytes were lysed into 0.5 ml of the same buffer. As the stability of the different activities varied, each cell line was lysed at the same time, centrifuged for 5 min at 10,000g, and the supernatant stored for 1 or 2 days at -20°C . Extracts from bovine brain and human placenta were prepared by homogenization and stored frozen, although these extracts were thawed several times before assay. Bovine brain cytosol was separated from membranes by centrifugation at 100,000g for 1 h. All lysates were assayed at a final concentration of 2 mg ml^{-1} extract protein in 0.2% NP-40, 20 mM HEPES, pH 7.5, 20 mM EDTA, 1 mM GTP, 2 mM DTT. Equal volumes of lysates and prebound N-ras.GTP (1 nM [γ - ^{32}P]GTP in 20 mM HEPES, pH 7.5, 2 mM MgCl_2 , 2 mM DTT) were incubated for 5 min at 22°C and assayed for phosphate release²³. Background hydrolysis was determined in the presence of 0.5 mg ml^{-1} Y13-259 (a rat monoclonal antibody²⁴ against p21^{ras} which blocks both GAP interactions³) and was subtracted from total hydrolysis. All values represent the average of several determinations, with standard deviations of less than 10%. Western blots probed with p120-GAP antibodies revealed that most of the extracts contained 0.5–2 ng of p120-GAP per 50 μg of cell protein, suggesting that p120-GAP concentrations are ~ 10 nM. As these extracts were assayed at a dilution of ~ 20 (~ 0.5 nM p120-GAP), the specific activities extrapolated from the above data are consistent with any dodecyl maltoside-insensitive GAP activity being due to p120-GAP. But immunoblots showed that HL 60 cells and PBLs contained significantly less p120-GAP, indicating that there may still be other p21^{ras} GAP activities.

to the GAPs. Also, inhibition of any particular GAP activity, such as lipid inhibition of NF1-GAP or protein kinase C-dependent inhibition of either GAP, should turn on the signals; again, it is implied that binding is unimpaired. Conversely, selective inactivation of a particular GAP signal (without impairing its GAP activity) should not affect the other GAP pathways. A therapeutic drug could possibly be designed to target a particular GAP signalling function. We are investigating the putative signalling functions of the two GAPs. □

Received 7 March; accepted 3 May 1991.

- Barbacid, M. A. *Rev. Biochem.* **56**, 779–827 (1987).
- McCormick, F. *Oncogene* **5**, 1281–1283 (1990).
- Martin, G. A. *et al. Cell* **63**, 843–849 (1990).
- Xu, G. *et al. Cell* **63**, 835–841 (1990).
- Ballester, R. *et al. Cell* **63**, 851–859 (1990).
- Trahey, M. *et al. Science* **242**, 1697–1700 (1988).
- Buchberg, A. M., Cleveland, L. S., Jenkins, N. A. & Copeland, N. G. *Nature* **347**, 291–294 (1990).
- Trahey, M. & McCormick, F. *Science* **238**, 542–545 (1987).
- Adari, H., Lowy, D. R., Willumsen, B. M., Der, C. J. & McCormick, F. *Science* **240**, 518–521 (1988).
- Calés, C., Hancock, J. F., Marshall, C. J. & Hall, A. *Nature* **332**, 548–551 (1988).

- Vogel, U. S. *et al. Nature* **335**, 90–93 (1988).
- Yatani, A. *et al. Cell* **61**, 769–776 (1990).
- Bourne, H. R., Sanders, D. A. & McCormick, F. *Nature* **349**, 117–127 (1991).
- Tsai, M.-H., Yu, C.-L., Wei, F. S. & Stacey, D. W. *Science* **243**, 522–526 (1989).
- Tsai, M.-H., Hall, A. & Stacey, D. W. *Molec. cell. Biol.* **9**, 5260–5264 (1989).
- Yu, C.-L., Tsai, M.-H. & Stacey, D. W. *Molec. cell. Biol.* **10**, 6683–6689 (1990).
- Downward, J., Graves, J. D., Warne, P. H., Rayter, S. & Cantrell, D. A. *Nature* **346**, 719–723 (1990).
- Goldschmidt-Clermont, P. J., Machesky, L. M., Baldassare, J. J. & Pollard, T. D. *Science* **247**, 1575–1578 (1990).
- Birchmeier, C., Broek, D. & Wigler, M. *Cell* **43**, 615–621 (1985).
- Tanaka, K. *et al. Molec. cell. Biol.* **10**, 4303–4313 (1990).
- Tanaka, K., Lin, B. K., Wood, D. R. & Tamanoi, F. *Proc. natn. Acad. Sci. U.S.A.* **88**, 468–472 (1991).
- Zhang, K. *et al. Nature* **346**, 754–756 (1990).
- Shacter, E. *Analyt. Biochem.* **138**, 416–420 (1984).
- Furth, M. E., Davis, L. J., Fleurdelys, B. & Scolnick, E. M. *J. Virol.* **43**, 294–304 (1982).

ACKNOWLEDGEMENTS. We thank R. White and D. Viskochil for NF1 DNA, R. Clark and W. Crosier for NF1-GRD, C. Der for H-ras plasmids, and M. Innis and P. Polakis for discussion. This work was supported by Hoffman-La Roche (F.M.) and by a postdoctoral fellowship from the NCI (G.B.).

Homology of a 150K cytoplasmic dynein-associated polypeptide with the *Drosophila* gene *Glued*

Erika L. F. Holzbaur, James A. Hammarback*, Bryce M. Paschal, Nancy G. Kravitz†, K. Kevin Pfister‡ & Richard B. Vallee

Worcester Foundation for Experimental Biology, Shrewsbury, Massachusetts, USA

† Department of Anatomy and Cell Biology, University of Virginia Health Sciences Center, Charlottesville, Virginia, USA

CYTOPLASMIC dynein is a microtubule-activated ATPase which produces force towards the minus ends of microtubules^{1,2}. It is thought to be responsible for retrograde axonal transport and other aspects of organelle motility^{2–6} and may have a role in the poleward movement of mitotic chromosomes^{7,8}. Cytoplasmic dynein is an oligomeric complex of two catalytic heavy chains and a number of accessory subunits^{1,9,10}. We now report the cloning and sequencing of a complementary DNA for one of these species, a cytoplasmic dynein-associated polypeptide of relative molecular mass 150,000 (M_r 150K). A full-length cDNA was found to contain an open reading frame of 4.0 kilobases, which is predicted to encode a polypeptide of M_r 145K. It has extensive homology with the product of the *Drosophila* gene *Glued*, which encodes a polypeptide of M_r 148K¹¹. The *Glued* mutation is dominant, with pleiotropic developmental defects in heterozygotes and an embryonic lethal phenotype in homozygotes. As dominant mutations may involve disruption of normal protein–protein interactions, the *Glued* mutation should provide insight into the mode of action of cytoplasmic dynein *in vivo*.

Brain cytoplasmic dynein, originally referred to as microtubule-associated protein 1C (MAP 1C), contains two subunits of M_r 410K (heavy chains), three of M_r 74K and about one each of M_r 59, 57, 55 and 53K^{1,9}. An additional polypeptide of M_r 150K was found in cytoplasmic dynein purified from other sources¹⁰. Whereas the heavy chains are thought to contain the sites of ATP hydrolysis and force production, the function of the other polypeptides is uncertain. In the case of *Chlamydomonas* flagellar dynein, at least one, and probably several, accessory subunits are associated with the base of the molecule¹². This suggests that the noncatalytic polypeptides play a regulatory role, perhaps in targeting the holoenzyme to different subcellular sites.

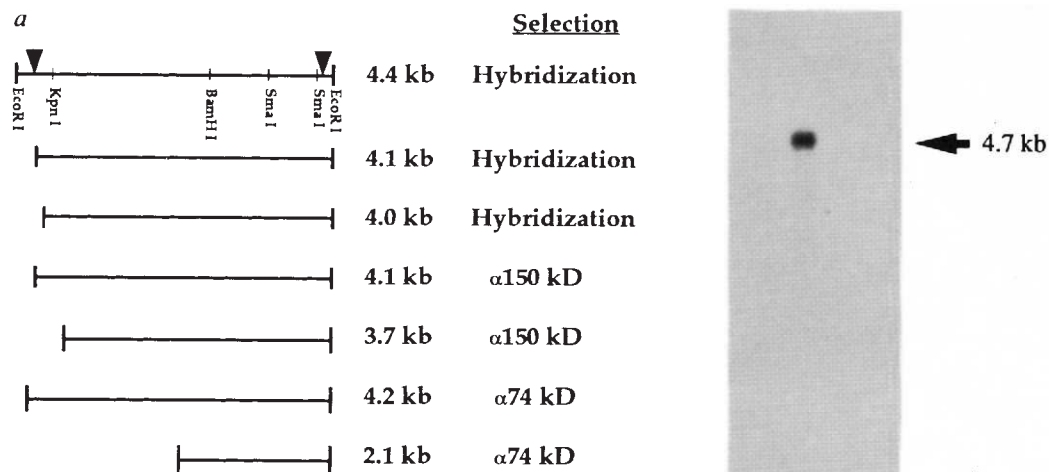
The 150K polypeptide has been found in cytoplasmic dynein preparations from a variety of sources^{5,8,10}. Bovine brain dynein contains only trace levels of the 150K polypeptide when isolated

Current addresses: * Department of Neurobiology and Anatomy (Molecular Genetics), The Bowman Gray School of Medicine, Wake Forest University, Winston-Salem, North Carolina, USA.

† Department of Genetics, Washington University School of Medicine, St. Louis, Missouri, USA.

FIG. 1 *a*, Restriction map of the full-length 150K polypeptide, with the sites of translation initiation and termination indicated by arrowheads. The overlapping clones are illustrated, in conjunction with the method of isolation, either immunoselection or hybridization. *b*, Northern blot analysis of the expression of the transcript encoding the 150K polypeptide in rat brain.

METHODS. The initial clones were selected by probing a size-selected rat brain λ ZAP cDNA expression library (>3 kb inserts). Immunoselection of clones was performed with either a mouse polyclonal anti-150K antiserum, prepared by immunization with bovine brain cytoplasmic dynein and preabsorbed by incubation with excess heat-stable MAPs²³ to eliminate minor apparent MAP 2 reactivity, or with a mouse monoclonal anti-74K antibody (K. K. Pfister, unpublished data). The library was then rescreened with nick-translated probes generated from *EcoRI*-*KpnI* or *EcoRI*-*BamHI* digests of the 4.2 kb



immunoselected cDNA insert to isolate the full-length clone. Northern blot analysis was performed on RNA isolated from 4-day postnatal rat brain. RNA isolation, electrophoresis, blotting and hybridization were performed according to standard methods²⁴. The nitrocellulose filter was probed with a 1.7 kb *BamHI*-*EcoRI* fragment from the 3' end of the cDNA insert.

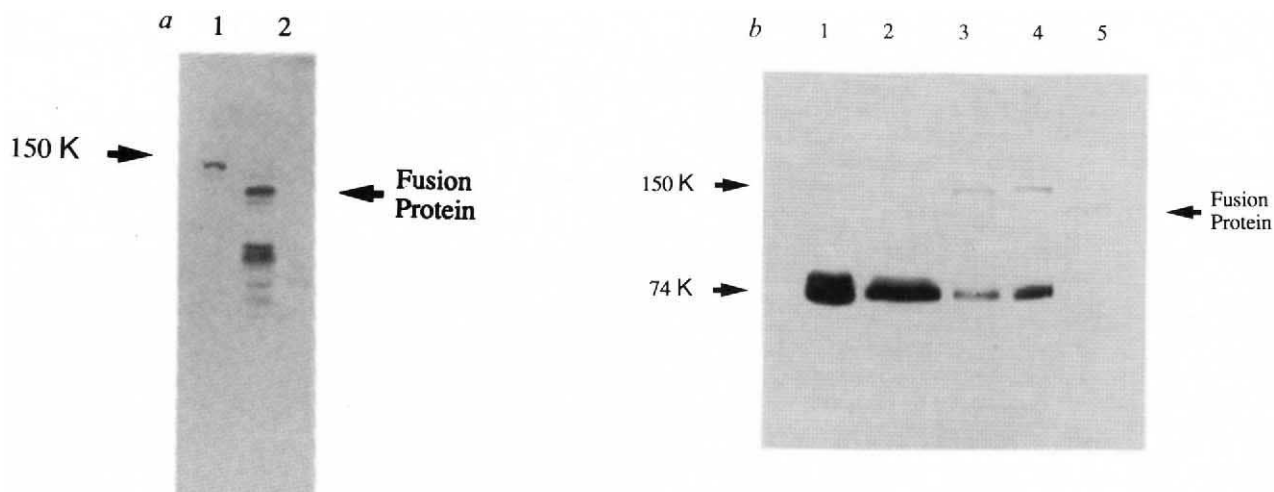


FIG. 2 Immunoblotting of purified bovine brain dynein and the 150K- β -galactosidase fusion protein with affinity-purified anti-150K mouse polyclonal antibody (*a*) and an anti-74K mouse monoclonal antibody (*b*). *a*, Lane 1, the 20S peak fraction from a sucrose gradient purification of NaCl-extracted bovine brain dynein¹³. Lane 2, the 150K polypeptide- β -galactosidase fusion protein in an *Escherichia coli* lysate. Multiple immunoreactive bands observed in the bacterially expressed fusion protein may be due to proteolysis. No reaction was observed with control lysed bacteria (data not shown). *b*, Lane 1, ATP extract of taxol-microtubules. Lane 2, the 20S peak fraction from a sucrose gradient purification of the sample shown in the previous lane. Lane 3, NaCl extract of taxol-microtubules. Lane 4, the 20S peak fraction from a sucrose gradient purification of the sample shown in the previous lane. Lane 5, 150K polypeptide- β -galactosidase fusion protein in an *E. coli* lysate.

METHODS. Cytoplasmic dynein was prepared from bovine brain according to the method of Paschal *et al.*¹, or by modification of this method including extraction of the microtubule with 0.4 M NaCl¹³. The β -galactosidase fusion protein was expressed from a construct of a 4.1 kb insert in frame in the pBluescript plasmid. The culture was induced for several hours with IPTG, then the cells were isolated, lysed, centrifuged briefly and mixed with SDS-sample buffer. The proteins were resolved by 7% SDS-PAGE, then electroblotted onto nitrocellulose, incubated with either affinity-purified anti-150K antibody which was visualized with alkaline phosphatase-conjugated secondary antibody (*a*), or with anti-74K antibody, which was then visualized with horseradish peroxidase-conjugated secondary antibody (*b*).

FIG. 3 Nucleic acid and predicted amino-acid sequences of the full-length 4.4 kb cDNA clone encoding the 150K dynein-associated polypeptide. Translation from the predicted start (270, overlined) to stop sites (4,245, overlined) would produce a polypeptide of 1,325 amino acids. The identity of the clone was verified by N-terminal microsequencing of two *Staphylococcus aureus* V8 proteolytic fragments of the purified bovine brain 150K polypeptide. The resulting peptides had N-terminal sequences of: LAFKASEQIYGTXS and STLKGAQMKASLAALP (underlined in the sequence; single-letter amino-acid notation).

METHODS. The cDNA inserts were excised from λ ZAP to yield the corresponding clones in pBluescript (Stratagene), which were sequenced on both strands by subcloning to produce *AccI*, *Clal*, *BamHI*, *BglII*, *HindIII*, *PstI*, *PvuII*, *SacI* and *SmaI* fragments; synthetic oligonucleotide primers were used to complete the sequencing. Double-stranded nucleotide sequencing was performed with the Sequenase kit (USB). Peptide sequencing was performed by purifying the 150K polypeptide from sucrose gradient-purified salt-extracted bovine brain dynein by electrophoresis through 7% SDS-PAGE. The polypeptide was then digested for 30 min in a second 12% gel with *Staphylococcus aureus* V8-protease²⁶. Following electroblotting to Immobilon-PVDF (Millipore), two peptides of 20 and 24K were excised and sequenced on an Applied Biosystems Model 470A sequenator.

1	GAATTC	CCCGCAGAGT	GAGACT	CAATCAGT	TGGG	GAGCGGGCT	CAGTGT	CTCTGGGTCAG	60												
61	GAATCCT	TGGCAGTAGA	AAGGACTT	ACTTCT	CAGCCCAT	CAGCTAGC	CTTACCCCTCCC		120												
121	GAGGGAG	GCCCATGGAT	TGAGGAGG	GTCCCT	TGCTGAC	GGCTTCT	GAGCCGGTCCATGCA	180													
181	GGCCCAT	CACCTGAC	CCCTCTCG	CTAGGTTT	CTGCTT	TCTGCTCT	TTCGGTAGCTG	240													
241	GGGAGGGGGT	AGAGTCCGGT	TGAGCAAGT	GGCCGAGAGCA	AGGAGGCACAT	GTACAATC		300													
				M	A	Q	S	K	R	H	M	Y	N	R	11						
301	GGAGCC	CCCACTG	GCGAGG	ATGAGT	ACCGAG	CAAGCCCG	CGCTCGGGT	CGCGTCC	360												
12	T	P	S	G	S	R	M	S	T	E	A	S	R	P	A	G	R	L	P	31	
361	CGTGTG	AGGATG	ATTTGG	AAGGCC	ACCGTGT	ACTGTC	TGCTAT	GTGGAG	CTACACT	420											
32	C	G	G	D	W	E	G	H	T	G	T	V	A	Y	V	G	A	T	L	F	51
421	TGGC	CACTGG	CAATG	GTGGTGG	CTGATCT	GATGAAG	CAAAAGG	CAAGAC	GTGTA	480											
52	A	T	G	K	W	V	G	V	I	L	D	E	A	K	G	K	N	D	G	T	71
481	CCGTTC	AGGCAAG	GAATTT	TTCATG	TCGATG	AAGGCC	ACGGCAT	CTTTG	TACGCC	540											
72	V	Q	G	R	K	Y	F	T	C	D	E	G	H	G	I	F	V	R	Q	S	91
541	CCGAC	CTCAAG	TAATTT	GAAGAT	TGAGCAG	TACTACT	TCCCC	AGAGCT	CTCTATT	600											
92	Q	I	Q	V	F	E	D	G	A	T	T	T	S	P	E	T	P	D	S	111	
601	CTGCTT	CAAGAT	CTCTCA	AGAGAC	AGGAGCT	GATGAC	GTGCA	AAAGAC	CAGCAAA	660											
112	A	S	K	I	L	K	R	E	G	A	D	A	A	A	K	T	S	K	L	R	131
661	GGGAG	CTGAAG	CTAAG	AAGC	CACCGAG	CCCAAG	CAACCA	ACTCG	ACGGCC	720											
132	G	L	K	P	K	K	A	P	T	A	R	K	T	T	T	R	R	G	P	A	151
721	CTACTG	CCAGC	AGCTACT	TGGGTGG	TGGCCAG	TAGCTC	CTTGGG	CCCTCT	GTGAT	780											
152	T	R	P	A	S	T	G	V	A	G	P	S	S	S	L	G	P	S	G	S	171
781	CAGCAT	CGCGCT	GAATCT	GAGCAG	CTGAGC	CCAGCC	AGCTCAG	ACTCCG	CTGG	840											
172	A	S	A	G	E	L	S	S	E	P	T	P	A	O	T	P	L	A		191	
841	CAGCACC	ATCAT	CCCCAC	ACGGCCCT	CACCTC	TCTTGG	AGCAGC	ACCCCACT	TCCTT	900											
192	A	P	T	I	P	T	P	A	L	T	S	G	A	A	P	P	L	P	S	211	
901	CTCCCT	TAAAGA	AGGAGG	CGCTGAGG	AGCAGGT	ACGGGAC	CTGAGG	AGAGG	AGCTGG	960											
212	P	S	K	E	E	E	G	L	R	D	Q	V	R	D	L	E	E	K	L	E	231
961	AGACCT	TGGCTT	GAAAGC	GTGAGA	AGCAAG	CAAGCT	GAAAGAG	CTGG	AAGCACA	1020											
232	T	L	R	L	K	R	S	E	D	K	A	K	L	K	E	L	E	K	H	K	251
1021	AGATCC	AGCTGG	AGCAGT	GCGAAGT	GGAGS	CAAACTG	CAGGAG	CAGCAG	CGGGAC	1080											
252	I	Q	L	E	Q	V	E	W	K	S	K	M	Q	E	Q	A	D	L		271	
1081	TGACGCG	CGCCTCA	AGGCAG	CGGCAAG	GAGCAAG	GAGCAG	CTGAGG	CAAAAG	GAGCGC	1140											
272	Q	R	R	L	K	A	G	E	G	S	Q	G	G	T	G	G	K	G	A	L	291
1141	TACATG	AGGAGAT	TGGCAG	ACACAG	CCGAGG	CTCATG	ATGCAAT	TGGCAG	CTGG	1200											
292	H	G	D	G	R	H	S	R	R	H	D	G	H	S	G	Q	G	D		1311	
1201	ATGCCG	AGGAGCG	CGGAGT	CCCTG	CAGCAGG	AGAGCT	TGAGGCA	TGAAGGA	AGAGTAG	1260											
312	G	R	G	A	A	E	S	L	Q	Q	E	V	E	A	L	K	E	R	V	D	331
1261	ACGAAC	TACCA	CAGACT	TGAGAT	CTCTA	AGGCTG	AAATG	TAGAGAGA	AAGCTCT	1320											

2381	632	CGCCGGGCTGCTTACTCTGGTGGATGTCGTCGAGCCACACTGCAACCGCTACGACGATG	2280
		A G L V Y Y S L S L L Q A T L H R Y E H A	671
2381	672	CCCTCTCTGATCGAGCTTGAGCTGTACAGCAAGCTGACAGCCCTGTACCCGACGATGA	2340
		L S Q C S Y D Y Y Y Y F V G G L T P C E M S	691
2381	712	CTGGCGAGGAGGCTCTTACGATTCTCTGACGAGCTCTGCTGACAGAGATGAGTGGAG	2400
		A H S E R S L D F L I E L L H E F D Q L D E	711
2401	712	AGACTCTCAACTCGACGCCCTCAGCAAGCGATTAAGACTACGACGACATCTGTACAGCA	2460
		T V N Y R E P L T F A I F Y Y Q H L T S I	731
2461	732	TCCACCTCGCTGAGCAGCCGCAAGAGCTGCACCATCCAGCTGGCTGACCATCAAGTTCA	2520
		H L A E Q P E E S T M Q L A D R I K F T	751
2521	752	CCGACGCTCGCTGGATCTGTATGAGCGTCGAGCTGGCCGCGCTGCTGCTCTTCTTCGAG	2580
		Q S A L I D D C M S V E Y V G R L P A F I Q	771
2581	772	CTGGCGCAGGCGCAACAGATATTGCTCTCTCTCTCCGAGACCTGGAAAGATCCTGCACTG	2640
		G Q E A T D I A I L L R D L E T S C S D	791
2641	792	ACATCCGTCAGTTCTGCAAGAGATCCGACGGCGAATGCCAGGACGGATGCTCTGGGA	2700
		I R Q F C K K I R R M R P G T D A P G I	811
2701	812	TCCGAGCAGCGCTGGCTCTGGCTCAGAGTGTCTGACACACTCCTGGACTCGAGGAAGC	2760
		P A A L A F G S Q Y S D T L L L D C R K H	831
2761	832	ACCTGACGTGGGTGCTACTCTTCTTCAGGAGGCTGGCAGCTCGACCCCGCAGCTCATCG	2820
		L T M V V A V L Q E V A A A A Q L I A	851
2821	852	CCCCCTTGCAGAGACGAGCGGCTGCCGTGCTGCATCGAGGAGCTGGCTTTCAAAG	2880
		P L A E N E G L P V A A L E E L A F K A	871
2881	872	CAAGCGGACGAGCTACCGGAGCCCTCCAGCACCCCTACGAATTCTCGCCGAGTCAT	2940
		S E Q I Y G S P S S S P Y E C L R Q S C	891
2941	892	GCTCCATCTCTCATCAGCAGATGAACAACTTGGCCAGCGCATCAGGAAGCGGATGATG	3000
		S I L I S T M N K L A T A M Q E G E Y D	911
3001	912	ACCGCAGCGGGCCCCGAGCAAGCTCTCTCAAGTTGAACCTTGGCTCGACAGCTCGGTG	3060
		A E R P P S K P P P V E P V P A A L R A	931
3061	932	CAGAGATCAGAGTGCTGAAGTCTTGGTTTGAAGCTTGAGGTCGAGAGACAGTCATCA	3120
		F I T T A D A E G L G L K L E D R E T V I K	951
3121	952	AGGAGTTAAGAAGTCACCTAAGAACTAAGGAGACGAGGACTGATCGGCGCAACCTCGGGC	3180
		E L K K S L K I K G E E L S E A N V R L	971
3181	972	TGAGCTCTCGGAGAGAAGCTGGACAGCGCTGCCAAGATCTGTATGAGCGAATCGAGA	3240
		S L L E K L D S A A K D A D E R I E K	991
3241	992	AAGTTTCAGACTCGCTGGAGGAGCGGACCTCTGATCGGAAAGGAGGAAGAGTTTG	3300
		V O T R I E E T O T T L L R K K E K E F E	1011
3301	1012	AGGAGACAATGGATGCACTCCAGCGTACATAGACCAGCTGGAGGCAGAGAAGACAGAGC	3360
		E T M D A L Q A D I D Q L E A E K T E L	1031
3361	1032	TCAAGCAGCGCTGAACAGCCAGTCCAAGCCCATATTGAGGGACTCCGGGGCCCCCTC	3420
		K Q R L N S Q S K R T I E G L R G P P P	1051
3421	1052	CATCAGGACTGTACCTCGTCTCTGGCATTTGCTGCTGAGGAACAACGCGAGGGGCA	3480
		S G I A T L V S G I A G E E Q Q R G G T	1071
3481	1072	CTCTGGGAGGCTCCAGCGCCCTTCCGAGCCGGGGCGGTGAGGAGCTCACCACCTGC	3540
		P G Q A P G A L L P G P G P V K D S P L R	1091
3541	1092	TGCTTCAGCAGATCTCTGCGATGAGGTTACACATCTCTCAGCTCCAGCATGAGAACAGCA	3600
		L Q Q I S A M R L H I S Q L Q H E N S I	1111
3601	1112	TCCTCGAGGAGCCGACGATGAAGCGCTCTTGGCAGCTCTCGCCCTCTGCATCTTGCAAC	3660
		L R G A Q M K A S L A A L P P L H V A K	1131
3661	1132	AGTTTTCCCTCCCAACCCTCAGGCGCTGGTGTAACTCTGTCTGGGCGACTGTATC	3720
		F S L P P H E G P G G N L L S G A L Y R	1151
3721	1152	CGAAGCAGCGACTGTGGAGAACTAAACCAACTGAGCAGCTACACCCACGATGAGT	3780
		K T S O L L E K L N L L S T Y T H V D	1171
3781	1172	ACATCACCCGAAGCAGCGCGGCTCGAAGAGCCCATCGCTCAGCTTATGGAAACAAGTG	3840
		I T R S S P A C K S P S A Q L M E Q V A	1191
3841	1192	CCGACCTCAAGTCTTGAGCGACACCATCGAAGAGCTCAAGGATGAGGTTCTCAAGGAGA	3900
		Q L K S L S D T I E K L K D V T H L K E T	1211
3901	1212	CAGTAACCTCAGCTCTGAGACTACTGTTCCCACTGACTTGGCCACCTTCCCTTCATCTG	3960
		V T Q R P P G A T V P T G A F A T F P S S A	1231
3961	1232	CCGCGCTGAGGCGCAAGGAAGAGCAGGAAGACGACACGATCTACATGGCAAGGTGACC	4020
		G R E G G G R A A R R H S L H G Q S D L	1251
4021	1252	TTCTGCTGTGGCGAGGCTCTGCAAGGAGACCGCTGCTGCTGACCCAGGAGCAGCTG	4080
		L V C G R P R T A T P P G A D P G A A A	1271
4081	1272	CACCGACTCAGGCTCTCATCTCTAAGCAGCTCTTCCCTGCTTCCCTTTGGTCCC	4140
		P A S R S S H L L S T S P S L P S L W S H	1291
4141	1292	ACTGCCCACTGACCGCGCACTCAGCCACTTCCGGGTGAACACATGCTAACTTTTGA	4200
		C P V H G H L S H S R V E P H A N F L S	1311
4201	1312	GTGGTCACTTCTGCTGTGTCGAGGTTCTGTGATCTCGTAACCTGAGTGGGTTTGGCT	4260
		G S P S P G L V R V L I L T P *	
4261		GTTTACCTTGGCTCTTCCATATTGCTGTTGATCTGCTGCTCTTCTTCTGAGGGGCT	4320
4321		CAGGAGATGAGGCGCTGACACCCCTCCACTACAGATTGACTGAGGTTCTCCTGACTGA	4380
4381		GGACTCACCCCTTGATTAAGCAACTTCTGCTTCAAAAAAAGGATTC	443

by ATP extraction of microtubules^{1,10}. But salt extraction of the microtubules released a subfraction of dynein containing the 150K species at stoichiometric amounts¹³. What appears to be the same polypeptide can be separated from the dynein holoenzyme by fast protein liquid chromatography; the resulting 150K-enriched fraction was reported to promote cytoplasmic dynein-mediated vesicle motility¹⁴. A monoclonal antibody to this polypeptide was also reported to label kinetochores, as with other anti-cytoplasmic dynein antibodies⁸.

To understand the role of the 150K polypeptide *in vivo* we obtained cDNAs encoding this species. A size-selected rat brain cDNA expression library constructed in λ ZAP was screened with a mouse polyclonal antibody raised against bovine brain cytoplasmic dynein (Fig. 2). Two clones with inserts of 4.1 and 3.7 kilobases (kb) were identified (Fig. 1). β -Galactosidase fusion proteins expressed by these clones reacted with antibody affinity-purified from the bovine brain 150K polypeptide (Fig. 2); conversely, antibody affinity-purified from the fusion proteins reacted specifically with the 150K dynein-associated polypeptide.

Ten more clones were selected from the same library using a monoclonal antibody to the 74K cytoplasmic dynein subunit. Nine of the 10 clones overlapped with the two clones selected using the anti-150K antibody (Fig. 1). This is due to cross-reactivity with the 150K polypeptide as indicated by immunoblotting (Fig. 2), suggesting that despite its differential biochemical behaviour the 150K polypeptide is structurally related to the 74K subunit.

The library was rescreened to obtain a full-length clone, resulting in five overlapping clones of 4.4 kb along with several partial clones of 4.1 and 4.0 kb (Fig. 1). Sequence analysis of the overlapping cDNAs revealed an open reading frame corresponding to 1,325 amino acids. Its consensus start site for translation initiation was AACATGG (ref. 15; Fig. 3) and translation is predicted to terminate at the stop codon TAA. The cDNA includes a poly(A) tail. The predicted M_r of the encoded polypeptide was 145K, consistent with the size of the bovine brain polypeptide observed by denaturing PAGE. The calculated pI of the protein was 6.4.

The bovine brain 150K polypeptide was exposed to *Staphylococcus* V-8 protease and fragments of 20 and 24K were subjected to N-terminal microsequencing (see legend to Fig. 3). The sequences of both peptides were found within the predicted amino-acid sequence obtained from the cDNA clones; minor differences presumably reflect evolutionary variation between the bovine and rat polypeptides (Fig. 3).

Southern blot analysis of bovine genomic DNA probed with a 0.8 kb *Bam*HI-*Sma*I fragment of the 150K polypeptide cDNA produced a restriction enzyme pattern consistent with the existence of a single gene (data not shown). Northern blot analysis revealed a single band at 4.7 kb, slightly larger than the longest cDNA obtained. The difference is likely to be due to the extent of polyadenylation of the transcript (Fig. 1b). No shorter messenger RNA species was observed when the northern blot was hybridized with either a partial or a full-length probe (Fig. 1b, and data not shown). This suggests that the 150K and 74K polypeptides are not generated by an alternative splicing mechanism but are probably products of distinct genes. The amino-acid sequence from the 74K bovine brain polypeptide was not found in the predicted 150K amino-acid sequence (B.M.P. and R.B.V., unpublished data), supporting this hypothesis. The observed antibody cross-reactivity (Fig. 2b) may instead reflect the occurrence of a common functional region in the 74K and 150K polypeptides, perhaps involved in the interaction with the dynein holoenzyme, with the organelle surface, or with mitotic kinetochores.

The Genbank, EMBL, Swissprot and NBRF databases were all searched with the nucleotide and predicted amino-acid sequences of the full-length cDNA. Substantial relatedness at both the nucleotide and protein sequence levels was found with

the *Drosophila* gene *Glued*, which encodes a polypeptide of 148K¹¹ (Fig. 4). Homology was observed throughout most of the length of the aligned polypeptides. Regions of highest sequence similarity were colinear throughout the two polypeptide sequences.

The original *Glued* mutation, *Gl*¹, is dominant, exhibiting pleiotropic developmental defects in heterozygotes, which are particularly apparent in aberrant eye and optic lobe formation¹⁶⁻¹⁸. Embryos homozygous for *Gl*¹ are not viable, nor are embryos homozygous for the null mutation. Genetic mosaic analyses failed to recover cells homozygous for *Gl*¹. Such a 'cell lethal' phenotype suggests a role for the gene in the viability and/or reproduction of the individual cell¹⁹. *In situ* hybridization studies showed that the *Glued* transcript is expressed in virtually all tissues of the embryo, late larva and pupa, suggesting a role in diverse cell types²⁰.

The dominant *Gl*¹ mutation is caused by a transposon insertion at the 3' end of the gene which results in the production of an abnormal truncated transcript with an altered C-terminal sequence²¹. The mutation seems to have no effect on the amounts of *Glued* transcript and seems not to perturb regulation of expression of the protein²¹. As the wild-type allele is haplo-sufficient, the mutant gene product was proposed to function in *trans* to affect the function of the wild-type polypeptide. Thus, the *Gl*¹ phenotype may involve abnormal protein-protein interactions between wild-type and mutant subunits of an oligomeric complex in which the mutant subunit poisons wild-type activity²¹. This hypothesis is supported by the observation that the effect of the heterozygous mutation can be titrated; that is supernumerary copies of the wild-type gene reduced the severity of the mutation¹⁹.

The 150K polypeptide may represent the mammalian homologue of this essential gene, or may be a member of a family of related genes. The identification of the *Glued* gene product as a cytoplasmic dynein subunit is consistent with the widespread tissue distribution of both proteins^{20,22}. As cyto-

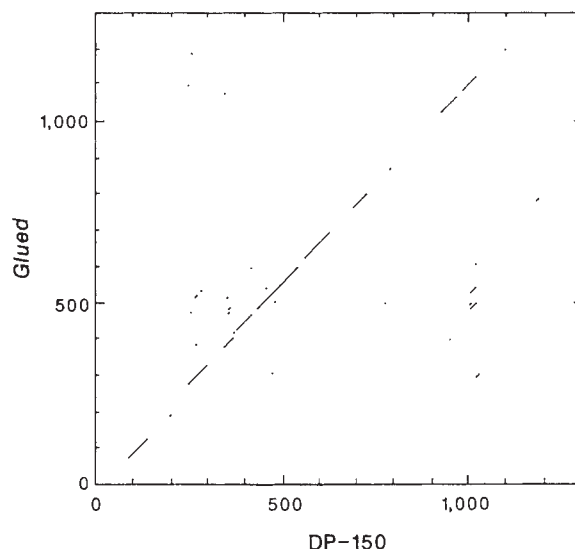


FIG. 4 Alignment of the predicted sequences of the 150K dynein-associated polypeptide (DP-150) and the *Glued* gene product (*Glued*) from *Drosophila* to yield a dot-plot comparison. The two sequences are 36% identical over the first 750 amino acids; a comparison of the entire two sequences shows 31% identity and 53% similarity, as determined by the GAP program²⁵.

METHODS. The databases GenBank, EMBL, Swissprot and NBRF were searched with both the nucleic acid and predicted protein sequences of the 150K polypeptide using the programs FASTA and TFasta²⁵. The *Glued* sequence was recognized at both the nucleotide and protein level. The dotplots shown were created with a stringency of 20 in a window size of 30 using the COMPARE and DOTPLOT programs²⁵. All sequence analyses were performed with the University of Wisconsin at Madison Genetic Computer Group Sequence Analysis Software Package (version 6.2)²⁵ as run on a VAX 11/750 computer.

plasmic dynein has been implicated in retrograde organelle transport and possibly in mitosis, it is likely that genetic alteration of a component of the dynein complex would result in a lethal disruption of cellular function, as was observed for *Glued*. Cytoplasmic dynein is particularly abundant in brain tissue, where it is thought to be responsible for retrograde axonal transport. This could explain the defective development of the eye and optic lobe in *Gl/Gl⁺* heterozygotes. The developmental relationship between the eye and the optic lobe is complex. But it has been shown from studies of genetic mosaics that *Gl/Gl⁺* neurons of the optic lobe develop abnormally even when innervated by wild-type photoreceptor cells¹⁸. Thus *Gl¹* has a direct effect on neuronal development. Whether this results from a defect in axonal transport as predicted for a cytoplasmic dynein mutant remains to be determined. □

Received 21 March; accepted 3 May 1991.

1. Paschal, B. M., Shpetner, H. S. & Vallee, R. B. *J. Cell Biol.* **105**, 1273–1282 (1987).
2. Paschal, B. M. & Vallee, R. B. *Nature* **330**, 181–183 (1987).
3. Forman, D. S., Brown, K. J. & Promesberger, M. E. *Brain Res.* **272**, 194–197 (1983).
4. Schnapp, B. J. & Reese, T. S. *Proc. natn. Acad. Sci. U.S.A.* **86**, 1548–1552 (1989).
5. Schroer, T. A., Steuer, E. R. & Sheetz, M. P. *Cell* **56**, 937–946 (1989).
6. Hirokawa, N. *et al. J. Cell Biol.* **111**, 1027–1037 (1990).
7. Pfarr, C. M. *et al. Nature* **345**, 263–265 (1990).
8. Steuer, E. *et al. Nature* **345**, 266–268 (1990).
9. Vallee, R. B. *et al. Nature* **332**, 561–563 (1988).
10. Collins, C. A. & Vallee, R. B. *Cell. Motil. Cytoskel.* **14**, 491–500 (1989).
11. Swaroop, A., Swaroop, M. & Garen, A. *Proc. natn. Acad. Sci. U.S.A.* **84**, 6501–6505 (1987).
12. King, S. M. & Witman, G. B. *J. Biol. Chem.* **265**, 19807–19811 (1990).
13. Holzbaur, E. L. F., Kravitz, N. G. & Vallee, R. B. *J. Cell Biol.* **109**, 157a (1988).
14. Schroer, T. A. & Sheetz, M. P. *J. Cell Biol.* **111**, 294a (1990).
15. Kozak, M. *Cell* **44**, 283–292 (1986).
16. Plough, H. H. & Ives, P. T. *Genetics* **20**, 42–69 (1935).
17. Meyerowitz, E. M. & Kankel, D. R. *Dev. Biol.* **62**, 112–142 (1978).
18. Garen, S. H. & Kankel, D. R. *Dev. Biol.* **96**, 445–466 (1983).
19. Harte, P. J. & Kankel, D. R. *Genetics* **101**, 477–501 (1982).
20. Swaroop, A. *et al. Molec. Cell Biol.* **6**, 833–841 (1986).
21. Swaroop, A., Paco-Larson, M. L. & Garen, A. *Proc. natn. Acad. Sci. U.S.A.* **82**, 1751–1755 (1985).
22. Yoshida, T., Takazumi, H. & Izutsu, K. *J. Cell Biol.* **111**, 296a (1990).
23. Vallee, R. B. *Meth. Enzym.* **134**, 89–104 (1986).
24. Sambrook, J., Fritsch, F. & Maniatis, T. *In Molecular Cloning* (Cold Spring Harbor Laboratory, New York, 1989).
25. Devereux, J., Haeberli, P. & Smithies, O. *Nucleic Acids Res.* **12**, 387–395 (1984).
26. Cleveland, D. W. *et al. J. Biol. Chem.* **252**, 1102–1106 (1977).

ACKNOWLEDGEMENTS. We thank D. Howland, S. Wadsworth and R. Obar for scientific advice and discussion. K. Angelides for the cDNA library, J. Leszyk for peptide sequencing and C. Wilkerson for assistance with sequence analysis software. The taxol used in this study was supplied by the Drug Synthesis and Chemistry Branch, Division of Cancer Treatment, National Cancer Institute.

teristic of GTP-binding proteins². Extensive homology was found between this domain and the mammalian Mx proteins which are involved in interferon-induced viral resistance^{3,4}, and with the product of the *VPS1* locus in *Saccharomyces cerevisiae*, which has been implicated both in membrane protein sorting⁵, and in meiotic spindle pole separation⁶. Dynamin-containing microtubule bundles were not observed in an immunofluorescence study of cultured mammalian cells⁷, but a role for a GTP-requiring protein in intermicrotubule sliding during mitosis in plants has been reported⁸. We report here that *Drosophila melanogaster* contains multiple tissue-specific and developmentally-regulated forms of dynamin, which are products of the *shibire* locus previously implicated in endocytic protein sorting^{9,10}.

An antibody prepared against rat brain dynamin identified a *Drosophila* protein with relative molecular mass of about 100,000 (M_r 100K) (Fig. 1a), which was most abundant in heads, less abundant in embryos, and was undetected in adult bodies. To evaluate the ability of the immunoreactive species to bind to microtubules, taxol was added to cytosolic extracts prepared

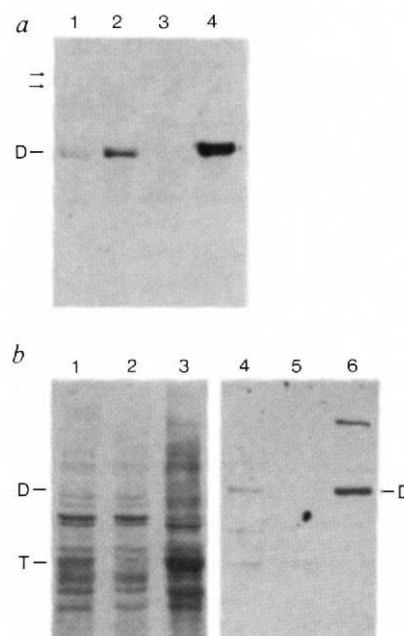


FIG. 1 Immunoblotting of dynamin in *Drosophila*. Antibody to a bacterially expressed glutathione S-transferase/rat brain dynamin fusion protein was reacted against *Drosophila* tissues and microtubule preparative fractions separated by SDS-PAGE (8% polyacrylamide¹⁷) and transferred to nitrocellulose¹⁸. The antibody does not cross-react with either bacterially expressed Mx1² or the product of the *VPS1* gene (C.C.S. and C. Vater, unpublished observations), and thus seems to be specific for dynamin. The 100K species was occasionally seen to split into two closely spaced bands. Additional embryonic species of 150–160K were consistently seen. *a*, Immunoblot of: lane 1, 0–2 h embryos; lane 2, heads; lane 3, bodies; lane 4, calf brain cytosolic extract. *b*, lanes 1 and 4, head cytosolic extract; lanes 2 and 5, supernatant and lanes 3 and 6, pellet from microtubule sedimentation, stained for protein using Ponceau S (lanes 1 to 3) or reacted with anti-dynamin antibody (lanes 4 to 6). D, dynamin, T, tubulin, arrows, high M_r immunoreactive species.

METHODS. For immunoblotting, wild-type *Drosophila* were frozen and agitated to separate heads from bodies. Antibody staining in *a* was developed using alkaline phosphatase-conjugated secondary antibody and in *b* was developed with the ECL Western Blotting Detection system (Amersham). For the preparation of microtubules¹⁹, heads were dissected from anaesthetized wild-type *Drosophila* adults, and were homogenized in a glass-on-glass homogenizer in ~5–10 volumes of 0.1 M Pipes, pH 6.6, containing 1 mM EGTA, 1 mM MgSO₄, 10 μ g ml⁻¹ leupeptin, 1 mM phenylmethylsulphonyl fluoride, and 10 μ g ml⁻¹ tosyl arginine methyl ester. The homogenate was centrifuged at 18,000 r.p.m. for 30 min at 4 °C to yield a cytosolic extract. After addition of taxol to a final concentration of 20 μ M the extract was incubated at 37 °C for 15 min and the microtubules were sedimented as above but at 37 °C.

Multiple forms of dynamin are encoded by *shibire*, a *Drosophila* gene involved in endocytosis

Maio S. Chen*, Robert A. Obar*, Christopher C. Schroeder*, Timothy W. Austin†, Clifton A. Poodry†, Samuel C. Wadsworth*‡ & Richard B. Vallee*‡

* Cell Biology Group, the Worcester Foundation for Experimental Biology, Shrewsbury, Massachusetts 01545, USA

† Department of Biology, University of California, Santa Cruz, California 95064, USA

DYNAMIN was discovered in bovine brain tissue as a nucleotide-sensitive microtubule-binding protein of relative molecular mass 100,000¹. It was found to cross-link microtubules into highly ordered bundles, and appeared to have a role in intermicrotubule sliding *in vitro*. Cloning and sequencing of rat brain dynamin complementary DNA identified an N-terminal region of about 300 amino acids which contained the three consensus elements charac-

‡ To whom correspondence should be addressed.

b

Ddyn3 PPMLPSRR*

Ddyn4 PPMLPSRVGAVGGAIQQSGANRYVPESMRGQVNQAVGQAAINELSNAFSSRFK*

Two full-length cDNA clones were isolated from a *Drosophila* head cDNA library¹¹ by hybridization with a rat dynamin cDNA clone². The Ddyn4 cDNA contains an open reading frame of 883 amino acids, compared with 851 amino acids for rat dynamin² (Fig. 2a). The Ddyn3 cDNA contains an open reading frame of 836 amino acids, differing from Ddyn4 only at the C terminus (Fig. 2b). Predicted M_r of 99K and 94K for the polypeptides encoded by Ddyn4 and Ddyn3, respectively, agree well with the electrophoretic mobilities of the major immunoreactive species in the head (Fig. 1).

tions are included it is similar over 81% of the sequence. The three N-terminal GTP-binding sequence motifs² are perfectly conserved between the rat and fly proteins (Fig. 2a). The highest level of sequence identity is in the roughly 300 amino-acid N-terminal GTP-binding domain (80%), the same region most highly conserved between dynamin, Vps1p/Spo15, and the Mx proteins. In contrast to Mx and Vps1p, however, the level of sequence identity between rat dynamin and the fly polypeptides remains high throughout most of the remainder of the respective sequences. The lowest amount of sequence identity is seen within the C-terminal 100 amino acids, but the unusually high proline content and basic pI observed in this domain in rat² are features also present in the two *Drosophila* sequences. Vps1p and Mx1 lack the basic, proline-rich C-terminal region, suggesting that it performs a function unique to dynamin.

FIG. 2 a, Alignment of deduced amino-acid sequence (single-letter amino-acid notation) of the Ddtn4 protein with rat dynamin². The sequences were aligned using the GAP program of the Genetics Computer Group with a GAP weight of 3.0 and a length weight of 0.1. The fly sequence is identical to 68% of the rat sequence and similar to about 81% when conservative amino-acid replacements are included. Consensus GTP-binding motifs are overlined. A consensus initiator methionine codon used by both cDNAs corresponds to Met 6 of the rat dynamin sequence and conforms well to the Kozak consensus sequence as modified for *Drosophila*^{20,21}. Presumptive terminator codons at the end of the Ddtn3 and Ddtn4 cDNA reading frames are followed by terminator codons in all reading frames. The arrowhead indicates the position after which Ddtn3 and Ddtn4 differ. The proline-rich C-terminal region previously noted in rat dynamin begins with Pro 753. The analogous regions in Ddtn3 and Ddtn4 contain 28 and 29 Pro residues, respectively, compared with 32 in rat. This domain is also characterized by a high ratio of basic to acidic residues; 2.5 in rat, 5 in Ddtn3, and 3.25 in Ddtn4. b, Alternative C termini of the deduced amino-acid sequence of Ddtn3 and Ddtn4. The sequence of Ddtn3 is identical to that of Ddtn4 up to Arg-835. Val 836 is replaced by an Arg codon in the Ddtn3 sequence and is immediately followed by a termination codon. Beyond the Arg 835 codon, the Ddtn3 and Ddtn4 cDNA sequences are unrelated.

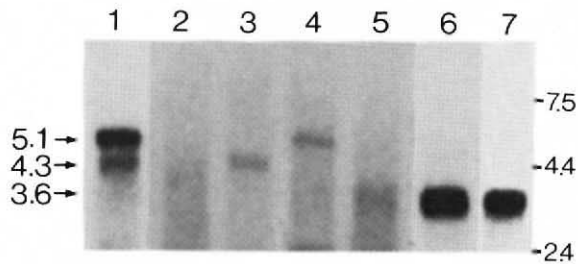


FIG. 3 *Drosophila* dynamin-related RNA species. Poly(A)-containing RNA (5–10 μ g) was hybridized with riboprobes for different regions of the Ddyn cDNAs. Probes recognizing both Ddyn3 and Ddyn4 cDNAs hybridized to two transcripts of 5.1 and 4.3 kb expressed in the head but not detected in the body (compare lanes 1 and 6). A Ddyn4-specific 3'-end probe recognized the 4.3 kb transcript (lane 3) whereas the Ddyn3-specific 3'-end probe recognized the 5.1 kb transcript (lane 4). An adult body-specific and an embryonic transcript of 3.6 kb also hybridized with the 5'-end probe common to both Ddyn3 and Ddyn4 cDNAs (lanes 6 and 7). The 3.6 kb body transcript is also faintly recognized by the Ddyn3 3'-specific probe (lane 5). Probes were as follows: I, a common probe from the 5'-untranslated region plus the first 21 nucleotides of Ddyn3 and Ddyn4; II, a common probe corresponding to the coding region of Ddyn4, excluding the first 600 nucleotides; III, a Ddyn3-specific 3'-untranslated probe; and IV, a Ddyn4-specific 3'-untranslated probe. Lane 1, heads probed with I and IV; lane 2, bodies probed with IV; lane 3, heads probed with IV; lane 4, heads probed with III; lane 5, bodies probed with III; lane 6, bodies probed with I, II and III; lane 7, embryos probed with I, II and III. Sizes of the dynamin-related RNA species are 5.1 kb, 4.3 kb and 3.6 kb. Sizes of RNA species used for calibration are indicated at right in kb.

METHODS. RNA was prepared from adult fly heads and bodies or embryos by extraction with guanidinium thiocyanate-phenol-chloroform²². Poly(A)-containing RNA was fractionated by oligo(dT) chromatography, electrophoresed on denaturing formaldehyde-agarose gels, and transferred to nitrocellulose. Riboprobe preparation and hybridization procedures have been previously published²³.

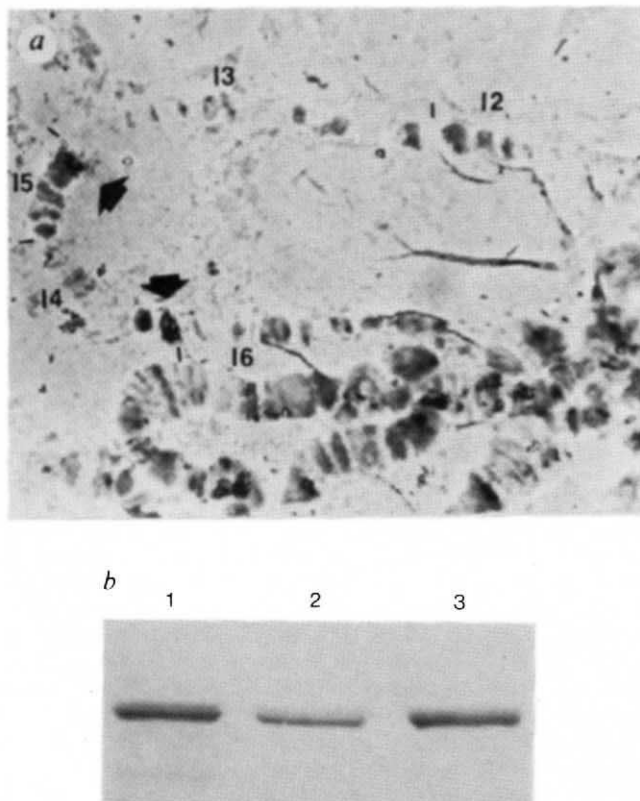


FIG. 4 *a*, Correlation of the *In(1)shi*¹⁶ breakpoint within the *shibire* gene with a cDNA clone from the region. The partial embryonic *shi-1* cDNA clone was hybridized with polytene chromosomes bearing the *In(1)shi*¹⁶ inversion breakpoint within the *shibire* gene. Sequences from the cDNA hybridize with both ends of the chromosomal inversion (arrows) indicating that this cDNA clone spans the inversion breakpoint. *b*, Correlation of the *In(1)shi*¹⁶ breakpoint within the *shibire* gene with the expression level of dynamin protein. Lanes 1 and 3, head protein from males bearing one functional copy of the *shibire* gene; lane 2, head protein from females bearing one functional copy of the *shibire* gene. Due to dosage compensation²⁴, males express each copy of an X chromosome gene at twice the relative level of an X chromosome gene in a female. Thus the individuals analysed in lanes 1 and 3 exhibit the wild-type expression level whereas the individuals analysed in lane 2 exhibit about 50% of the wild-type expression level. *c*, Correlation of the *shibire* cDNA with dynamin cDNA. The portion of the *shi-1* cDNA overlapping the

shi-1 RQREHSCKEQILLIDFELAYMNTNHEDFIGFANAQNKSENANKTGTQRQL
Ddyn3 461 RQREHSCKEQILLIDFELAYMNTNHEDFIGFANAQNKSENANKTGTQRQL
GNQVIRKGHMVIQNLGIMKGGSRPYFVLTSESISWYKDEDEKEKKFMLP
511 GNQVIRKGHMVIQNLGIMKGGSRPYFVLTSESISWYKDEDEKEKKFMLP
LDGLKLRDIEQGFMSMSRRVTFALFSPDGRNVYDYKQLELSCETVEDVE
561 LDGLKLRDIEQGFMSMSRRVTFALFSPDGRNVYDYKQLELSCETVEDVE
SWKASFLRAGVYPEKQETQENGDE-----SASEESSDPQLERQVETIR
611 SWKASFLRAGVYPEKQETQENGDEEGQEQKSASEESSDPQLERQVETIR
NLVDSYMKIVTKTTRDMVPKAIMMLIINNADKFINGELLAHLAYASGDQQAQ
661 NLVDSYMKIVTKTTRDMVPKAIMMLIINNADKFINGELLAHLAYASGDQQAQ
MMEESAESATRREEMLRMYRACKDALQIIGDVSMATVSSPLPPPVKNVDWL
711 MMEESAESATRREEMLRMYRACKDALQIIGDVSMATVSSPLPPPVKNVDWL
PSGLDNPRLSPPSPGCVGRKPGPPAQSSLGGRNPPLPPTGTPAPAIPINR
761 PSGLDNPRLSPPSPGCVGRKPGPPAQSSLGGRNPPLPPTGTPAPAIPINR
PGGGAPPLPGGRPGGSLPPPMLPSRR*
811 PGGGAPPLPGGRPGGSLPPPMLPSRR*

Ddyn3 cDNA was sequenced. With the exceptions noted in the text and indicated in the alignment, the *shi-1* cDNA sequence matches exactly with a position corresponding to amino acid 461 in Ddyn3 and extending to the end of the Ddyn3 cDNA. Based on these sequence comparisons, the *shibire* cDNA corresponds to the Ddyn3 class of mRNA (FIG. 2b).

METHODS. *a*, The *shi-1* cDNA clone was digoxigenin-labelled and hybridized with polytene chromosomes from *In(1)shi*¹⁶/*Dp(1); Yshi*⁺Y # 1 individuals. (The Y-borne duplications of the *shibire* region are described in reference 25.) *b*, All lanes contain protein from equal numbers of heads prepared and immunoblotted with alkaline phosphatase-conjugated secondary antibody as described in the legend to Fig. 1. The genotypes of flies analysed were: lane 1, *Df(1)sd*^{72b}/*Dp(1); Yshi*⁺Y # 3 (males); lane 2, *In(1)shi*¹⁶/*In(1)shi*¹⁶/*Dp(1); Yshi*⁺Y # 1 (females); lane 3, *In(1)shi*¹⁶/*Dp(1); Yshi*⁺Y # 1 (males). (*Df(1)sd*^{72b} deletes genetic material between 13F1 and 14B1²⁶, including the *shibire* locus.) *c*, The *shi-1* cDNA was sequenced and compared with the Ddyn3 open reading frame as indicated in the legend to Fig. 2.

head isoforms are encoded by distinct messenger RNA species differing in their 3' ends (Fig. 3). These experiments also revealed yet another transcript of 3.6 kilobases (kb) not detected in the head, and enriched in the adult body and embryo.

Results of genomic Southern blot and polytene chromosome hybridization were consistent with the existence of a single gene encoding the dynamin-related polypeptides, located within the 13F–14A region of the X chromosome (data not shown). Immunoblots of heads obtained from flies carrying duplications or deficiencies within the 13F–14A interval further localized the dynamin gene to the 14A region (data not shown) in which the *shibire* gene is known to be located^{12,13}. Temperature-sensitive *shibire* mutations cause reversible paralysis and defects in the endocytic pathway in neurons and other cell types^{9,10}. Dynamin is enriched in neuronal tissue (Fig. 1)¹⁷, and the dynamin-related yeast gene, *VPS1*, is itself involved in membrane sorting⁵, suggesting that dynamin could be the product of the *shibire* gene. The *shibire* locus has been cloned by chromosomal walking through the 14A region and the *shibire* gene identified within the cloned DNA segments by mapping the positions of chromosomal breakpoints known to disrupt *shibire* function (*In(1)shi*¹⁶, termed *In(1)shi*^{8–18} in reference 14, and *T(1;3)shi*¹⁹)^{13,14} (T. W. Austin, S. Park and C. A. Poodry, manuscript in preparation). A 2.3 kb partial embryonic cDNA clone (*shi-1*) was isolated by hybridization with the genomic DNA segment spanning the breakpoints noted above. Sequences within the *shi-1* cDNA were found to span the *In(1)shi*¹⁶ and *T(1;3)shi*¹⁹ breakpoints (Fig. 4a and data not shown). Results of two types of experiments indicate that the dynamin proteins are encoded by the *shibire* gene. First, we find that flies heterozygous for the *In(1)shi*¹⁶ mutation express about half of the wild-type amount of dynamin protein (Fig. 4b). Second, the deduced amino-acid sequence of the *shi-1* cDNA (Fig. 4c) exactly matches that of the Ddyn3 open reading frame from amino-acid position 461 to the terminator codon with two exceptions; the substitution of a lysine in the *shibire* sequence for Arg 598 in the Ddyn4 sequence, and the absence of six amino acids in the *shibire* sequence corresponding to amino acids 635–640 in the Ddyn4 sequence. The latter observation suggests that the *shi-1* embryonic cDNA represents yet another alternatively spliced form of mRNA from the dynamin/*shibire* gene.

Our results show that the *shibire* gene encodes the *Drosophila* equivalent of dynamin indicating a role for this protein in endocytosis. The paralytic phenotype is thought to result from a failure in synaptic vesicle reformation, reflected in a decrease in synaptic vesicles at the neuromuscular junction and an increase in coated or 'collared' pits^{9,10}. Aberrant membranous structures were also seen within the cytoplasm, but defects in later stages of the endocytic pathway were not specifically investigated. Microtubules have a role in translocation of endocytic vesicles from the periphery of the cell toward the centre^{15,16}, and some aspects of the *shibire* phenotype may reflect disruption of this process. As microtubules are not known to play a direct part in synaptic vesicle recycling, the effects at the synaptic terminal may reflect a backup of intermediates involved in the later stages of endocytosis. Alternatively, the few microtubules in the synaptic region may be more important for membrane budding than previously appreciated. A third possibility is that dynamin is multifunctional. In this regard, the dual membrane sorting and meiotic spindle pole separation defects observed in yeast *VPS1*/*SPO15* mutants are relevant. Hopefully, as a result of the findings of this study, the full range of dynamin functions will now be accessible to exploration *in vivo*. □

6. Yeh, E., Driscoll, R., Coltrera, M., Olins, A. & Bloom, K. *Nature* **349**, 713–715 (1991).
7. Scaife, R. & Margolis, R. L. *J. Cell Biol.* **111**, 3023–3033 (1990).
8. Asada, T., Sonobe, S. & Shibaoka, H. *Nature* **350**, 238–241 (1991).
9. Poodry, C. A. & Edgar, L. *J. Cell Biol.* **81**, 520–527 (1979).
10. Kosada, T. & Ikeda, K. *J. Cell Biol.* **97**, 499–507 (1983).
11. Itoh, N. et al. *Proc. natn. Acad. Sci. U.S.A.* **83**, 4081–4085 (1986).
12. Grigliatti, T. A., Hall, L., Rosenbluth, R. & Suzuki, D. T. *Molec. gen. Genet.* **120**, 107–114 (1973).
13. Poodry, C. A. *Drosophila Information Service* **68**, 207–208 (1990).
14. Poodry, C. A. *Dev. Biol.* **138**, 464–472 (1990).
15. Herman, B. & Albertini, D. F. *J. Cell Biol.* **98**, 565–576 (1984).
16. Matteoni, R. & Kreis, T. E. *J. Cell Biol.* **105**, 1253–1265 (1987).
17. Laemmli, U. K. *Nature* **227**, 680–685 (1970).
18. Bloom, G. S. & Vallee, R. B. *J. Cell Biol.* **96**, 1523–1531 (1983).
19. Vallee, R. B. *Meth. Enzym.* **134**, 104–116 (1986).
20. Kozak, M. *Nucleic Acids Res.* **15**, 8125–8148 (1987).
21. Cavener, D. R. *Nucleic Acids Res.* **15**, 1353–1361 (1987).
22. Chomczynski, P. & Sacchi, N. *Analyt. Biochem.* **162**, 156–159 (1987).
23. Gregory, R. J., Kammermeyer, K. L., Vincent, W. S. & Wadsworth, S. C. *Molec. cell. Biol.* **7**, 2119–2127 (1987).
24. Lucchesi, J. C. A. *Rev. Genet.* **7**, 225–237 (1973).
25. Poodry, C. A. *Drosophila Information Service* **55**, 210 (1980).
26. Cramer, L. & Roy, E. *Drosophila Information Service* **55**, 200–204 (1980).

ACKNOWLEDGEMENTS. We thank L. Ohn for technical assistance, A. Katzen, A. Chovnick and K. Matthews (IU Stock Center) for contribution of fly stocks, P. Salvaterra for the head library and N. Kravitz for helpful discussions. The taxol used in this study was supplied by the Drug Synthesis and Chemistry Branch, Division of Cancer Treatment, National Cancer Institute. This work was supported by the NIH, the National Cancer Institute and the National Science Foundation.

Large differences in the helix propensities of alanine and glycine

Avijit Chakrabarty*, John A. Schellman† & Robert L. Baldwin*

* Department of Biochemistry, Stanford University School of Medicine, Stanford, California 94305, USA

† Institute of Molecular Biology, University of Oregon, Eugene, Oregon 97403, USA

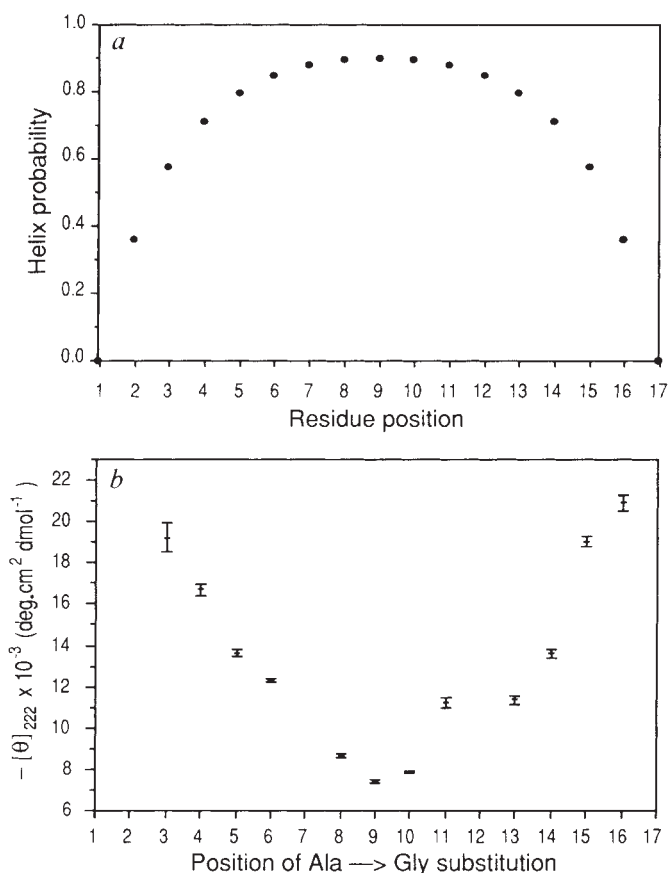
THE standard view of α helix formation in water, based on helix propensities determined by the host-guest method^{1,2}, is that differences in helix propensity among the amino acids are small, except for proline³, and that the average value of the helix propagation parameter s is near 1. A contradictory view of α helix formation in water is emerging from substitution experiments with short, unique-sequence peptides that contain only naturally occurring amino acids^{4–9}. Short peptides that contain only alanine and lysine, or alanine and glutamate, form surprisingly stable monomeric helices in water⁹ and substitution of a single alanine residue by another amino acid in these or related peptides produces a wide range of changes in helix content, depending on which amino acid is substituted for alanine^{4–6,8}. We show here that the ratio of the helix propensities of alanine to glycine is large, about 100, in substitution experiments with a 17-residue reference peptide containing alanine and lysine. The helix propensity is identified with s , the helix propagation parameter of the statistical mechanics model for α helix formation, and the results are interpreted by the Lifson–Roig theory¹⁰. Single alanine→glycine substitutions have been made at a series of positions in individual peptides. The helix-destabilizing effect of an Ala→Gly substitution depends strongly on its position in the helix, as predicted by the Lifson–Roig theory if the ratio of s values for Ala:Gly is large.

Amino-acid substitution experiments, in which changes in α helix formation by short peptides are measured, yield nearly the same rank order of helix propensities for different types of reference peptides used^{4–6,8}, but a different rank order from that obtained by host-guest experiments^{1,2}. The change in helix content for a substitution (Ala→X) nevertheless depends markedly on the choice of reference peptide in short-peptide experiments. Therefore, it is important to determine quantitative helix propensities by determining values of s , the helix propagation parameter of the Zimm–Bragg model¹¹, and to find out if the values of s vary with the choice of reference peptide. It has

Received 2 April; accepted 21 May 1991.

1. Shpetner, H. S. & Vallee, R. B. *Cell* **59**, 421–432 (1989).
2. Obar, R., Collins, C. A., Hammarback, J. A., Shpetner, H. S. & Vallee, R. B. *Nature* **347**, 256–261 (1990).
3. Staeheli, P., Haller, O., Boll, W., Lindenmann, J. & Weissmann, C. *Cell* **44**, 147–158 (1986).
4. Staeheli, P. & Sutcliffe, J. G. *Molec. cell. Biol.* **8**, 4524–4528 (1988).
5. Rothman, J. H., Raymond, C. K., Gilbert, T., O'Hara, P. J. & Stevens, T. H. *Cell* **61**, 1063–1074 (1990).

FIG. 1 *a*, Fractional helicity at each residue in the reference peptide computed from Lifson-Roig theory¹⁰. The computation employed an average s value of 1.56 and a σ value of 0.0029. The average value of s was derived by fitting the experimental helix content of the reference peptide, using the value of σ determined by J. M. Scholtz *et al.* (manuscript in preparation). The ordinate gives the computed helicity at a specified residue position. The Lifson-Roig¹⁰ theory of the α helix-random coil transition gives the partition function directly as a matrix product using the formula $Z = v_i \cdot \Pi M_j \cdot v_r$, where ΠM_j is an ordered sequence of Lifson-Roig matrices M_j set up to correspond with the amino-acid sequence; v_i and v_r are the usual end vectors¹⁰. The matrices contain growth and nucleation parameters w and v^2 which correspond to the s and σ of the Zimm-Bragg theory¹¹; s and σ are computed from w and v^2 by the relations $s = w/(1 + v)$ and $\sigma = v^2/(1 + v)^4$ (H. Qian and J.A.S., manuscript in preparation). Individual values of w and v^2 may be assigned for each amino acid in the sequence. J. M. Scholtz *et al.* (manuscript in preparation) have tested the applicability of the theory by measuring thermal unfolding curves for peptides of varying chain length (6–51 residues) that contain repeating sequences (AAKAA or AEAAGA in one-letter amino-acid code). A tentative value for the helix nucleation parameter, $v^2 = 0.0037$ or $\sigma = 0.0029$, has been obtained in these experiments. Once the partition function has been set up, the helical probability for the j th amino acid in the sequence can be obtained by differentiating $\ln Z$ with respect to $\ln w_j$, where w_j is the growth parameter for the j th amino acid in the sequence. Summing over all residues gives the mean number of helical units per chain, from which the fractional helicity can be calculated. For discussion of the molecular nature of s and σ , see for example ref. 18, and for a discussion of helix-coil transition theory, see ref. 19. *b*, α -Helical contents of the glycine-substituted peptides. The data points show the mean and the standard deviation ($n=5$). The experimental conditions are listed in Table 1.



been observed⁹ that the helix contents of alanine peptides, solubilized by insertion of a few lysine residues, are too large to be explained by the s values found using the host-guest method^{1,2}. A possible explanation has been given⁹, based on the special helix-forming properties¹² of the host residue (hydroxybutyl- or hydroxypropyl-L-glutamine), in host-guest experiments. The Lifson-Roig theory¹⁰ is used to interpret our results because it combines simplicity with an accurate representation of α helix cooperativity, and because it is well suited to the calculation of helicity at each residue position. The correct representation of cooperativity is particularly important in short helices because end effects are prominent. The results have, however, been converted to the more familiar s and σ parameters of the Zimm-Bragg theory. For details of the calculations by the Lifson-Roig theory, see the legend to Fig. 1.

The rationale of our experiments is as follows. Using $\sigma = 0.0029$, one finds that the end residues of short, partially helical

helices should be strongly frayed; see Fig. 1*a*. Consequently, the effect of substituting a helix-destabilizing residue will be larger at the centre of the helix than near either end. Consider the substitution Ala → Gly, because substitution experiments⁴⁻⁸ show that Ala is one of the best helix-forming residues and Gly is one of the most helix-destabilizing residues. If the ratio $s(\text{Ala}) : s(\text{Gly})$ is large, the position-dependent effect of an Ala → Gly substitution will also be large, according to prediction based on the Lifson-Roig theory. Thus, a series of single Ala → Gly substitutions was made at different residue positions in individual peptides, using the same reference peptide (sequences are given in Table 1), and the change in helix content was measured for each substitution. The peptides contain tyrosine as the N-terminal residue in order to determine peptide concentration accurately from tyrosine absorbance⁹. The α -NH₂ and α -COOH groups are blocked to avoid unfavourable charge interactions with the helix dipole¹³.

FIG. 2 Experimental and computed fractional helix contents. See Table 1 for conditions. The data points and the curves represent experimental and computed values, respectively. Experimental fractional helix contents were determined from $[\theta]_{222}$ measurements using $-40,000(1 - 2.5/n)$ and $0 \text{ deg cm}^2 \text{ dmol}^{-1}$ as the values for 100 and 0% helix, respectively; n is the number of amino-acid residues in the peptide²⁰, which has an additional peptide group at each end because of the acetyl and amide blocking groups. The curves were computed from the Lifson-Roig equation using an average s value for Ala, Lys, and Tyr of 1.56, $s(\text{Gly}) = 0.015, 0.08, 0.20$, and 0.50 , and $\sigma = 0.0029$.

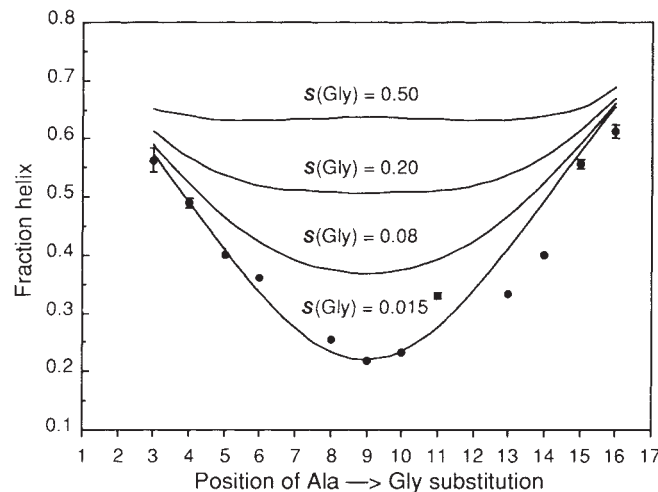


TABLE 1 Sequences and helix contents of the reference and glycine-substituted peptides

Name	Sequence	$[\theta]_{222}^*$ (deg cm ² dmol ⁻¹)
Reference	CH ₃ CONH-YKAAAAKAAAAKAAAK-CONH ₂	-25,000 ± 300
Gly 3	CH ₃ CONH-YKGAAGKAAAAKAAAK-CONH ₂	-19,200 ± 700
Gly 4	CH ₃ CONH-YKAGAAKAAAAKAAAK-CONH ₂	-16,700 ± 300
Gly 5	CH ₃ CONH-YKAAAGKAAAAKAAAK-CONH ₂	-13,600 ± 200
Gly 6	CH ₃ CONH-YKAAAGKAAAAKAAAK-CONH ₂	-12,300 ± 100
Gly 8	CH ₃ CONH-YKAAAAKGAAGKAAAAK-CONH ₂	-8,700 ± 100
Gly 9	CH ₃ CONH-YKAAAAKAGAAKAAAAK-CONH ₂	-7,400 ± 100
Gly 10	CH ₃ CONH-YKAAAAKAGAAKAAAAK-CONH ₂	-7,800 ± 100
Gly 11	CH ₃ CONH-YKAAAAKAAAGKAAAAK-CONH ₂	-11,200 ± 200
Gly 13	CH ₃ CONH-YKAAAAKAAAGKAAAK-CONH ₂	-11,400 ± 200
Gly 14	CH ₃ CONH-YKAAAAKAAAGKAAAK-CONH ₂	-13,600 ± 200
Gly 15	CH ₃ CONH-YKAAAAKAAAGKAAAGK-CONH ₂	-19,000 ± 300
Gly 16	CH ₃ CONH-YKAAAAKAAAGKAAAGK-CONH ₂	-20,900 ± 400

Peptides were synthesized by the solid-phase method using the simultaneous multiple peptide synthesis procedure¹⁷. An active ester coupling procedure, employing pentafluorophenyl esters of 9-fluorenylmethoxycarbonyl amino acids, was used. The amino termini were acetylated with acetic anhydride and the peptides were cleaved from the PAL resin (Milligen) as peptide-amides with 95:5 trifluoroacetic acid/anisole mixture. The peptides were purified by C18 reverse-phase chromatography, and peptide identity was confirmed by FAB mass spectrometry. Peptide purity was assessed by capillary electrophoresis, and ranged between 94–100% purity (average purity, 98.6%).

* Mean residue ellipticity at 222 nm. Conditions: 0 °C, pH 7.00, 1.0 M NaCl, 1 mM each of sodium borate, sodium citrate, and sodium phosphate, and 10–30 μM peptide. Measurements were made on an Aviv 60DS spectro-polarimeter in a 1.0-cm path-length cuvette. The values represent the mean ± s.d. (n=5). There is an additional error of ±1% associated with each measurement which was due to the error of measuring peptide concentration by tyrosine absorbance.

The results (Fig. 1b) show that the position effect is large for Ala → Gly substitutions. The largest drop in helix content is produced at residue 9, in the centre of the helix, and substitutions close to either end have comparatively small effects. The shape of the position-effect curve in Fig. 1b is inverse to that of the predicted curve of helix fraying (Fig. 1a). Figure 2 shows the data fitted to the Lifson–Roig theory for various values of $s(\text{Gly})$: a very low value (about $s = 0.015$) is required to fit the data. An average value of $s = 1.56$ fits the helix content of the reference peptide. As the reference peptide contains chiefly alanine, the ratio of $s(\text{Ala})$: $s(\text{Gly})$ is about 100. A few data points deviate significantly from the theoretical curve. Possibly small but abrupt changes in the properties of the alanine helix arise at the position where each Lys⁺ residue is inserted. As measured by the host-guest method¹, the nucleation constant σ for glycine is reported to have the small value of 10^{-5} . We studied the effect of allowing $\sigma(\text{Gly})$ to be 10^{-5} on the predicted curve of peptide helicity versus position of a Ala → Gly substitution. The predicted effect is quite small and cannot account for the experimental results. Likewise, allowing $s(\text{Lys})$ to be 5-fold smaller than $s(\text{Ala})$ has only a small effect on the predicted curve of helicity versus position of an Ala → Gly substitution, when the average s for Ala, Lys and Tyr remains the same for the reference peptide.

Two effects have been noted that should contribute to the low value of $s(\text{Gly})$. The greater flexibility of the glycine peptide backbone (see ref. 14, and references therein) probably favours the random coil (but see ref. 15) and an enthalpic interaction between Cβ and the α helix backbone favours the helix¹⁶ and is missing in glycine.

To compare directly our results for Ala → Gly substitutions with other short peptide results, it will be necessary to determine the s values of Ala and Gly in the other reference peptides. Our peptide is similar to the 17-residue alanine-based peptide used by Merutka *et al.*⁵, although their peptide contains three possible helix-stabilizing ($i, i+4$) E⁻...K⁺ salt bridges, and the change we observe for an Ala → Gly substitution at a central position in the helix is similar to the change they observe. A considerably smaller change was found by Lyu *et al.*⁶: using a 21-residue peptide to make three central Ala → Gly substitutions, they found a change in helix content that is comparable to the change found

here for a single Ala → Gly substitution. Their reference peptide contains eight possible ($i, i+4$) E⁻...K⁺ salt bridges, whereas our reference peptide has no known helix-stabilizing side-chain interaction of importance. O'Neil and De Grado⁸ studied an Ala → Gly substitution on the solvent-exposed face of a dimeric coiled-coil helix, and interpreted the result by the two-state equation. More work is needed to relate their result to ours.

In summary, the large ratio found here for $s(\text{Ala})$: $s(\text{Gly})$ indicates that the helix propensities of the amino acids are widely distributed, in contradiction to host-guest results. The existence of this contradiction suggests that context-dependent, or neighbouring residue, effects need to be analysed in future to understand fully the nature of substitution experiments. A strong dependence of the effect of a substitution on position in the helix has been demonstrated for the Ala → Gly substitution, and it has been explained by use of the Lifson–Roig equation. □

Received 27 December 1990; accepted 29 April 1991.

1. Sueki, M. *et al.* *Macromolecules* **17**, 148–155 (1984).
2. Scheraga, H. A. *Proc. natn. Acad. Sci. U.S.A.* **82**, 5585–5587 (1985).
3. Altmann, K.-H., Wójcik, J., Vasquez, M. & Scheraga, H. A. *Biopolymers* **30**, 107–120 (1990).
4. Padmanabhan, S., Marqusee, S., Ridgeway, T., Laue, T. M. & Baldwin, R. L. *Nature* **344**, 268–270 (1990).
5. Merutka, G., Lipton, W., Shalongo, W., Park, S.-H. & Stellwagen, E. *Biochemistry* **29**, 7511–7515 (1990).
6. Lyu, P. C., Liff, M. I., Marky, L. A. & Kallenbach, N. R. *Science* **250**, 669–673 (1990).
7. Strehlow, K. G. & Baldwin, R. L. *Biochemistry* **28**, 2130–2133 (1989).
8. O'Neil, K. T. & DeGrado, W. F. *Science* **250**, 646–651 (1990).
9. Marqusee, S., Robbins, V. H. & Baldwin, R. L. *Proc. natn. Acad. Sci. U.S.A.* **86**, 5286–5290 (1989).
10. Lifson, S. & Roig, A. *J. chem. Phys.* **34**, 1963–1974 (1961).
11. Zimm, B. H. & Bragg, J. K. *J. chem. Phys.* **31**, 526–535 (1959).
12. Lotan, N., Yaron, A. & Berger, A. *Bipolymers* **4**, 365–368 (1966).
13. Shoemaker, K. R., Kim, P. S., York, E. J., Stewart, J. M. & Baldwin, R. L. *Nature* **326**, 563–567 (1987).
14. Creighton, T. E. *Proteins* 162–170 (Freeman, New York, 1983).
15. Brant, D. A., Miller, W. G. & Flory, P. J. *J. molec. Biol.* **23**, 47–65 (1967).
16. Gö, M., Gö, N. & Scheraga, H. A. *J. chem. Phys.* **54**, 4489–4503 (1971).
17. Houghten, R. A., DeGraw, S. T., Bray, M. K., Hoffman, S. R. & Frizzell, N. D. *Biotechniques* **4**, 522–528 (1986).
18. Cantor, C. R. & Schimmel, P. R. *Biophysical Chemistry* Vol. III (Freeman, New York, 1980).
19. Poland, D. & Scheraga, H. A. *Theory of Helix–Coil Transitions in Biopolymers* (Academic, New York, 1970).
20. Chen, Y.-H., Yang, J.-T. & Chau, K. H. *Biochemistry* **13**, 3350–3359 (1974).

ACKNOWLEDGEMENTS. We thank H. Qian and J. M. Scholtz for their help, the Mass Spectrometry Facility, University of California, San Francisco (supported by an NIH grant) for verification of peptide identity by mass spectrometry, and the Protein and Nucleic Acid Facility, Beckman Center, Stanford University School of Medicine, for measurement of peptide purity by capillary zone electrophoresis. This work was supported by a grant from the US National Science Foundation to R.L.B. and from the US National Institutes of Health to J.A.S. A.C. is a fellow of the Medical Research Council of Canada.

Reduced binding of TFIID to transcriptionally compromised mutants of VP16

C. James Ingles, Michael Shales, W. Douglas Cress*, Steven J. Triezenberg* & Jack Greenblatt

Banting and Best Department of Medical Research, University of Toronto, Toronto, Ontario M5G 1L6, Canada

* Department of Biochemistry, Michigan State University, East Lansing, Michigan 48824, USA

ACTIVATOR proteins that control transcription initiation by RNA polymerase II^{1,2} usually have two domains: one binds to DNA, and the other activates transcription^{3,4}. A particularly potent acidic^{5,6} activation domain at the C terminus of the herpes simplex virus protein VP16^{7–9} binds directly and selectively to the human and yeast TATA box-binding factor TFIID¹⁰. We have now investigated the biological significance of this *in vitro* interaction by using mutant forms of VP16¹¹. For changes at the critical phenylalanine residue at position 442 of VP16 there was a good correlation between transactivation activity *in vivo* and the binding of VP16 to TFIID *in vitro*. In contrast, mutants with reduced negative charge were more defective for binding than for activation.

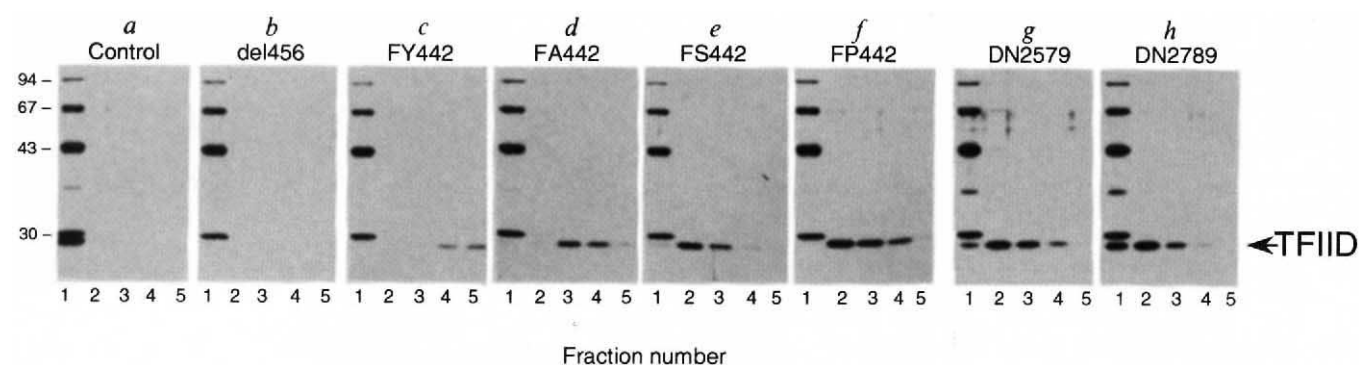


FIG. 1 Affinity chromatography with the yeast TFIID polypeptide detects the effects of missense mutations in the activation domain of VP16. The purified yeast TFIID polypeptide¹⁰ of relative molecular mass 27,000 (M_r , 27K) (2 μ g) was mixed with marker proteins (Biorad) and chromatographed¹⁰ on 20 μ l pA-VP16 columns, or a 20 μ l control column (a) containing no bound protein. The columns were washed with 50 μ l aliquots of loading buffer (5–10 min per aliquot). Each fraction (50 μ l) therefore represented elution with 2.5 column volumes of loading buffer. Aliquots (37 μ l) of the flow-through (lane 1) and successive wash fractions (lanes 2–5) were analysed on 12.5% SDS-polyacrylamide gels which were silver stained¹⁵. When TFIID bound to the column it was absent in the flow-through and wash fractions (for example b) or retarded (for example c) relative to the molecular mass markers. With

the exception of b and c where TFIID interacted strongly with pA-VP16, the recoveries of TFIID in the flow-through and wash fractions was 70–80%. The differences in chromatographic behaviour of TFIID on the various VP16 columns (for example g versus h) have been reproducible (2–5 times). For the expression of VP16^{del456} and other mutant derivatives of VP16, *Sac1* to *HindIII* fragments encoding amino acids 423–456 of the activation domain of each VP16 derivative in plasmid pMSVP16del456¹¹ were subcloned between the VP16 *Sac1* site and the *Sma1* site of the pA-VP16 expression plasmid¹⁰. pA-VP16 polypeptides were purified from *E. coli* extracts and coupled¹⁰ to Affigel 10 (Biorad) at a concentration of 1.5 mg ml⁻¹ (b–f) or 3.0 mg ml⁻¹ (g, h).

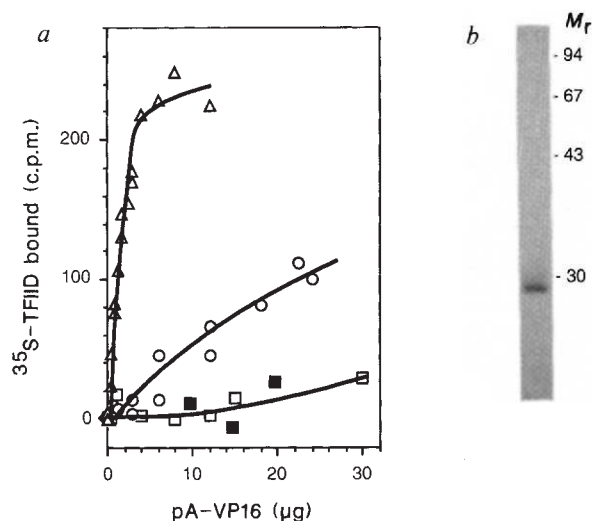
In our original affinity chromatography experiments we used the complete C-terminal activation region (amino acids 413–490) of VP16 expressed as a fusion protein with *Staphylococcus aureus* protein A (pA) (ref. 10). Because the effects *in vivo* of point mutations in VP16 were studied in the context of a C-terminal truncation of its activation domain (del456, with amino acids 456–490 deleted)^{7,11}, we produced the del456 activation domain and its mutant derivatives as pA-VP16 fusions in *Escherichia coli*. Purified yeast TFIID was mixed with a set of marker proteins and loaded onto affinity columns containing purified pA-VP16^{del456} fusion polypeptides coupled covalently to agarose. By monitoring the flow-through and wash fractions for the TFIID polypeptide, we assessed the strength of the VP16-TFIID interaction. TFIID was found in the flow-through fraction (fraction 1) from control columns containing no immobilized protein (Fig. 1a) or immobilized pA (ref. 10). As with the full-length activation domain of VP16 (ref. 10), however, yeast TFIID bound to a pA-VP16^{del456} column (Fig. 1b) and could only be eluted by a high-salt buffer (ref. 10, data not shown).

Amino acid 442, a phenylalanine residue in wild-type VP16,

is particularly important for transcriptional activation^{11,12}. VP16^{del456} derivatives in which this Phe 442 is changed to Pro (FP442), Ala (FA442) or Ser (FS442) failed to activate transcription *in vivo*¹¹, or *in vitro* in the case of FP442¹². Yeast TFIID bound all three mutant VP16 derivatives poorly *in vitro* (Fig. 1d, e, f). With a more conservative change from Phe to Tyr at position 442, VP16 retained about 35% of its activity *in vivo*¹¹, and pA-VP16^{FY442} bound TFIID much better (Fig. 1c) than did pA-VP16 derivatives with other changes at position 442; TFIID only started to elute after a wash of 10 column volumes. Thus, for changes at position 442, transactivation potential *in vivo* correlated well with the strength of the interaction between yeast TFIID and VP16 as monitored by affinity chromatography. This implies that the interaction between the acidic activation domain of VP16 and the basic (pI = 9.8) yeast TFIID polypeptide is not simply ionic, and instead involves some degree of stereochemical specificity. Either the Phe at position 442 makes a critical hydrophobic contact with TFIID, or changes at this position disrupt the structure of the VP16-activation domain and indirectly alter the surface that binds to TFIID.

The strength of the VP16-TFIID interaction was measured

FIG. 2 a, Binding of ³⁵S-TFIID to pA-VP16 is reduced by mutations in VP16. *E. coli* BL21(DE3) cells¹⁶ expressing yeast TFIID from the bacteriophage T7 promoter in the plasmid pAR3038 (ref. 17) were grown in 50 ml M9 medium containing 10⁻⁴ M Na₂SO₄ and 400 μ M H₂[³⁵S]O₄. ³⁵S-labelled TFIID was purified as previously described¹⁰. ³⁵S-labelled TFIID (12 ng; 0.5 Ci mol⁻¹) was incubated for 8 min at 25 °C in 50 μ l binding buffer (10 mM HEPES, pH 7.9, 100 mM NaCl, 1 mM EDTA, 20% glycerol, 1 mg ml⁻¹ BSA) with various amounts of pA-VP16 (Δ), pA-VP16^{del456} ($\circ$), pA-VP16^{FP442} ($\square$), or pA (ref. 10) ($\blacksquare$). After a further 10 min incubation with 100 μ l 1:1 suspension of IgG-Sepharose 6 Fast Flow beads (Pharmacia) in binding buffer, the beads were pelleted (2 min, 300g), washed with 900 μ l binding buffer, and counted in CytoScint (ICN). A background of 80 c.p.m. obtained in the absence of pA-VP16 was subtracted. b, Radioautograph of the SDS gel-fractionated ³⁵S-labelled TFIID preparation. Positions of marker proteins shown on right ($M_r \times 10^{-3}$).



by a modification of the affinity method. ^{35}S -labelled yeast TFIID (Fig. 2b) was incubated in solution with varying concentrations of pA-VP16. IgG-Sepharose beads were then added to bind protein A and thereby sequester the pA-VP16-TFIID complexes. The wild-type full-length pA-VP16 bound TFIID much better than did pA-VP16^{del456} (Fig. 2a). Binding to pA-VP16^{FP442} was barely detectable and not distinguishable from binding to protein A itself. TFIID bound to the wild-type full-length VP16 activation domain (amino acids 413-490) in pA-VP16 with a dissociation constant (K_d) of $2 \times 10^{-7} \text{ M}^{-1}$ in buffer containing 0.1 M NaCl. The del456 derivative bound TFIID 30 times less well ($K_d = 6 \times 10^{-6} \text{ M}^{-1}$). The presence of the inactivating FP442 mutation further reduced the affinity of VP16 for TFIID by at least 10-fold.

We also examined the behaviour of two VP16 mutants, DN2579 and DN2789, in each of which four of the acidic residues were changed to Asn or Gln¹¹. DN2579 had 60% the activity *in vivo* of the del456 parent whereas the DN2789 derivative had only 18% activity¹¹. In this class of mutations, the correlation between activities *in vivo* and binding of TFIID *in vitro* also held (Fig. 1g, h). TFIID bound to pA-VP16^{DN2579} (peak in fractions 2 and 3) better than it did to pA-VP16^{DN2789} (peak in fractions 1 and 2). But both pA-VP16^{DN2579} and pA-VP16^{DN2789} bound yeast TFIID less well than the transcriptionally inactive Phe 442 mutants. Thus, when comparing different classes of VP16 mutations, the correlation between activation *in vivo* and binding *in vitro* did not hold. Although acidic residues are important for activation by VP16 *in vivo*¹¹, the ionic component of the VP16-TFIID interaction may be more important in our *in vitro*-binding experiments using yeast TFIID than it is for activation of transcription in mouse cells.

The effects of mutations at Phe442 establish a correlation between transactivation activity *in vivo* and the binding of VP16 to TFIID *in vitro*. This suggests that TFIID is a target for the acidic activation domains of transcription factors such as VP16, and that the strength of binding of activators to TFIID is an important parameter governing the rate of transcription initiation in eukaryotes. This activator-TFIID interaction may be only one aspect of transcriptional activation. That there is a marked cooperativity in transcriptional activation by multiple DNA-bound activator proteins¹³ suggests there may be a second target for activators in the basic transcription machinery. Similar experiments have demonstrated that at higher salt and ligand concentrations VP16 binds directly or indirectly to the general initiation factor TFIIB¹⁴. Multiple activator proteins may accelerate the initiation of transcription by bringing together TFIID and another general initiation factor like TFIIB. Our identification of TFIID as one of the targets for acidic activators is an important step in understanding the molecular mechanisms of transcriptional activation. $\square$

Received 4 March; accepted 19 April 1991.

1. Mitchell, P. & Tjian, R. *Science* **245**, 371-378 (1989).
2. Johnson, P. F. & McKnight, S. L. *A. Rev. Biochem.* **58**, 799-839 (1989).
3. Brent, R. & Ptashne, M. *Cell* **43**, 729-736 (1985).
4. Ptashne, M. *Nature* **335**, 683-689 (1988).
5. Hope, I. A. & Struhl, K. *Cell* **46**, 885-894 (1986).
6. Ma, J. & Ptashne, M. *Cell* **48**, 847-853 (1987).
7. Triezenberg, S. J., Kingsbury, R. C. & McKnight, S. L. *Genes Dev.* **2**, 718-729 (1988).
8. Sadowski, I., Ma, J., Triezenberg, S. & Ptashne, M. *Nature* **335**, 563-564 (1988).
9. Cousens, D. J., Greaves, R., Goding, C. R. & O'Hare, P. *EMBO J.* **8**, 2337-2342 (1989).
10. Stringer, K. F., Ingles, C. J. & Greenblatt, J. *Nature* **345**, 783-786 (1990).
11. Cress, W. D. & Triezenberg, S. J. *Science* **251**, 87-90 (1991).
12. Berger, S. L., Cress, W. D., Cress, A., Triezenberg, S. J. & Guarante, L. *Cell* **61**, 1199-1208 (1990).
13. Carey, M., Lin, Y.-S., Green, M. R. & Ptashne, M. *Nature* **345**, 361-364 (1990).
14. Lin, Y.-S., & Green, M. R. *Cell* **64**, 971-981 (1991).
15. Morrissey, J. H. *Analyt. Biochem.* **117**, 307-310 (1981).
16. Studier, F. W. & Moffat, B. A. *J. molec. Biol.* **189**, 113-130 (1986).
17. Rosenberg, A. H. *et al. Gene* **56**, 125-135 (1987).

ACKNOWLEDGEMENTS. We thank L. Demirjian for technical assistance, K. Stringer and D. Fitzpatrick for discussion and M. Green for communication of results before publication. This work was supported by the MRC (Canada) and the National Cancer Institute of Canada (C.J.I. and J.G.) and the NIH and the Michigan Agricultural Experiment Station (S.J.T.). W.D.C. was supported by a Barnett Rosenberg graduate fellowship.

nature
is available in
microform.



University Microfilms

International reproduces this publication in microform: microfiche and 16mm or 35mm film. For information about this publication or any of the more than 13,000 titles we offer, complete and mail the coupon to: University Microfilms International, 300 N. Zeeb Road, Ann Arbor, MI 48106. Call us toll-free for an immediate response: 800-521-3044. Or call collect in Michigan, Alaska and Hawaii: 313-761-4700.

☐ Please send information about these titles:

Name _____

Company/Institution _____

Address _____

City _____

State _____ Zip _____

Phone () _____

University
Microfilms
International